AN ARTIFICIAL INTELLIGENT GRAIN QUALITY EVALUATION SYSTEM FOR INDIAN GRADING STANDARD

Synopsis of the PhD Thesis

Submitted by

Sachin Anil Sonawane

(Roll no. P021)

Department of Electronics and Telecommunication Engineering

Supervisor

Dr. Suresh Kurumbanshi

submitted in partial fulfilment of the requirements of

DOCTOR OF PHILOSOPHY (Engineering)

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING
SVKM's NARSEE MONJEE INSTITUTE OF MANAGEMENT STUDIES (Declared as Deemed-to-be University Under Section 3 of the UGC Act, 1956)
V. L. Mehta Road, Vile Parle (West)
MUMBAI – 400056

June 2025

TABLE OF CONTENTS

TABLE OF CONTENTS 2

ABSTRACT 4

Chapter 1 5

INTRODUCTION 5

Chapter 2 7

LITERATURE REVIEW 7

2.1 Problem Definition 9

2.2 Objectives 9

Chapter 3 10

LINEAR PATH INTERPOLATION FOR OBJECT DETECTION 10

3.1 Computational Flow for the Proposed Image Processing Scheme 10

3.2.1 Sampling Harvested Soybeans 10

3.2.2 Two-Sided Color Image Acquisition 11

3.2.3 Thresholding to Binarize Acquired Images 12

3.2.4 Removal of Dust Particles 13

3.2.5 Linear Path Interpolation (LPI) 14

3.2.6 Object Detection 23

3.2.7 Feature Extraction and Classification of Harvested Seeds 23

Chapter 4 24

HYBRID NEURAL NETWORK TO CLASSIFY HARVESTED BEANS 24

4.2.1 Sampling Harvested Black-Eyed Peas 24

4.2.2 Color Image Acquisition 25

4.2.3 Object Detection 26

4.2.4 Image Augmentation 26

4.2.5 Dataset Split 27

4.2.6 Object Classification using EfficientNetV2B1+ViT Model 27

Chapter 5 29

RESULTS AND DISCUSSION 29

5.1 Results of the Proposed Classification Approach Based on LPI Segmentation for Harvested Soybeans 29
5.2.1 Separation of Physically Touching Objects using LPI 30
5.2.2 Classification of Objects Detected using LPI 32
5.2 Results of the Proposed Hybrid Neural Network for Classifying Black-Eyed Peas 34
5.3.1 Sample Seed Detection using SAM 34
5.3.2 Classification of Detected Sample Seeds using Hybrid Neural Network 35
5.3.3 Performance Comparison Between the Existing and Proposed Methods 37
Chapter 6 38
SUMMARY AND CONCLUSION 38
Chapter 7 39
FUTURE WORK 39
PUBLICATIONS 40
REFERENCES 41

ABSTRACT

Harvested beans, such as soybeans and black-eyed peas, are important commercial crops, as their products, including green diesel, biomaterial products, animal food, and oil, are widely used in day-to-day life. The quality of harvested beans is essential to the food industry to ensure the standard of these products. Inspecting the physical properties of harvested beans is a preliminary step in quality evaluation. The conventional manual inspection of the physical properties of harvested beans is erroneous. Therefore, automated systems are preferred over manual inspection. This research addresses the challenge of improving accuracy in harvested bean classification based on the physical properties of the beans, focusing on improving the preservation of bean information during image capturing and bean detection that affects classification accuracy.

In the existing literature, the bean detection operations generally followed the watershed algorithm to identify individual beans from a clustered view of a harvested bean sample. Nevertheless, the effectiveness of the watershed algorithm is inadequate for harvested beans as it is susceptible to variations in the color features of the sample objects. A novel computing scheme integrating two-sided image acquisition and the proposed Linear Path Interpolation (LPI) method is introduced in this work. The proposed LPI method utilized topological features of the physically touching objects from a clustered view and the pattern-matching windows to detect the topological features. Experimental findings revealed that the proposed LPI method could perform bean detection with an accuracy of 97%. Also, the proposed method separates the physically touching objects at the contact region only, which improves the preservation of bean information in the bean disconnection process. The detected beans were classified using a two-stage model of weighted KNN and Logistic regression. Its classification accuracy was 90% for non-cracked sound-quality soybeans. The classification accuracy achieved for cracked and split soybeans varied between 70% and 90%, for all other categories it was above 90%.

Despite these achievements, the proposed LPI method observed limitations in a few cases where the sample image included randomly shaped organic foreign objects, such as sticks and stems. However, such cases are usually fewer for a random harvested bean sample.

Further, to improve the system's capability to identify individual harvested beans from clustered sample images, an advanced instance segmentation algorithm, the Segment Anything Model (SAM), integrated with custom object retrieval, was implemented. SAM was fed sample images captured under varying lighting conditions and different sample sizes. Furthermore, five hybrid deep learning models consisting of MobileNetV2 with ShuffleNetV2, EfficientNetV2B1 paired with Vision Transformer (ViT), InceptionV3 with MobileNetV2, ResNet50 with DenseNet121, and VGG16 integrated with ResNet18 were experimented for classification of detected objects. These models employed local and global features through advanced attention mechanisms. The model of EfficientNetV2B1 and ViT attained the maximum overall classification accuracy of 95.80%.

Keywords: Harvested beans, Image processing, Instance segmentation, Bean detection, Transfer learning with hybrid model, classification

Chapter 1

INTRODUCTION

Beans are among the most versatile agricultural crops worldwide, offering substantial nutritional, economic, and environmental benefits. Beans constitute a dietary staple for billions of people, providing an excellent source of plant-based protein, carbohydrates, and nutrients. Dry beans and pulses supply approximately 10% of the global population's protein intake and 5% of dietary energy. In low-income countries, this contribution may rise up to 20% of protein intake [1-3].

The versatility of beans has led to their use in many processed food items and applications. The processed bean products include vegetable oil, bean flour, canned beans, bean-based snacks, livestock feed, biofuels, bioplastics, products like tofu, tempeh, textured vegetable protein, soy sauce, etc. This overview explores the diverse applications of commonly harvested beans such as soybeans, black-eyed peas, and kidney beans in our daily lives [1-3].

However, the overall grade of these products is closely tied to the quality of the harvested beans used in their production. Therefore, the food industries must assess the quality of harvested beans as ultimately their products influence consumers' health. The food industries assess the quality of harvested beans and seeds as determined by the following criteria.

- Botanical properties (like species/ variety)
- Physical properties
- Chemical Characteristics

The current research aims to assess the physical properties of harvested beans or seeds. This assessment is conducted during the initial quality check through a visual method. For this purpose, the international or national grading standards [4,5] identified by the assessors are referenced. Typically, each country has its grading standards. Figure 1.1 shows the grading standard for harvested soybeans in India, as prescribed by Agmark, India, with the physical properties of harvested soybeans highlighted in a red square box.

In the past, the process of visual inspection was performed manually by human inspectors. However, this process is prone to inconsistency due to variations in human capability, often leading to significant judgment errors. Such errors may generate heavy financial burdens for soybean producers and the food industries that depend on such farm produce. To avoid these financial losses, it is necessary to implement an independent automated system devoid of human intervention. The non-destructive nature of automated systems ensures that physical damage to the beans is prevented.

The automated systems for the quality evaluation of harvested beans utilize image processing techniques and machine-learning models to detect and classify the beans with higher accuracy. These systems begin with image acquisition under controlled environmental conditions like lighting, imaging setup, etc., followed by image processing steps of noise reduction, image normalization, and image segmentation to detect individual beans. Then, feature extraction

identifies distinctive characteristics of color distributions, texture patterns, surface defects, and morphological parameters such as size, shape, circularity, etc. These extracted features serve as inputs to machine learning models that can differentiate between sound-quality beans and anomalies. The precise design of these components within automatic systems enhances the reliability of physical quality evaluations in bean processing, efficiently resolving the shortcomings of traditional manual inspections.

I) Agmark specifications:

The Agmark grade standards for Soybean notified under the Agricultural Produce (Grading and Marking) Act 1937 by the Central Government (Directorate of Marketing and Inspection) are given in Table No.- 4.

Table No.–4

Grade designation and definition of quality of Soybean seeds

Grade desig-nations	Special requirements							
	Oil content on dry basis percent by weight	Acid value of oil	Moisture content percent by weight	Damaged, discoloured, insect infested beans percent by weight	Immature, shrivelled beans percent by weight	Splits, brokens, cracked beans percent by weight	Inorganic foreign matter percent by weight	Organic foreign matter percent by weight
	Mini-mum	Maxi-mum	Maxi-mum	Maxi-mum	Maxi-mum	Maxi-mum	Maxi-mum	Maxi-mum
Grade –I	20	3	10	1	2	5	0.5	0.5
Grade – II	18	4	12	2	3	10	0.5	0.5
Grade –III	15	6	12	3	5	20	0.5	1.5

Figure 1.1: Indian grading standard for soybean by Agmark, India (Courtesy: https://agmarknet.gov.in/Others/soybean-profile.pdf)

This work focuses on improving image acquisition, image segmentation, and classification blocks to elevate the performance of the classification system of automated systems. The design of the image acquisition system aims to capture maximum information about every object in the harvested bean sample by acquiring two-sided images of the samples. The steady position of the bean sample and the controlled environment of the image acquisition setup ensure coordinate matching between the two views of respective objects in the harvested bean sample. Furthermore, the advancements in the image segmentation block aim to detect individual objects from the acquired sample images with improved preservation of bean information. For this purpose, the present work proposed a novel morphological image processing-based Linear Path Interpolation (LPI) algorithm. Similarly, to improve classification performance, this work carefully selected, designed and optimized key factors such as feature extraction, data augmentation, and optimization of hyperparameters of classification models. For this purpose, the present work experimented with extracting manually selected features using traditional image processing-based methods and feature extraction using deep neural networks. To enhance the model's ability to generalize, the training dataset was diversified using different data augmentation techniques. Additionally, the present work experimented with classification using basic machine learning algorithms and advanced pre-trained deep neural networks.

Chapter 2

LITERATURE REVIEW

The physical quality assessment of harvested seeds/ beans has undergone various changes due to the integration of advanced computer vision and machine learning techniques. In the past, people evaluated seeds through manual inspection to check for size, shape, color, and anomalies in harvested seeds. Now, automated systems with advanced technologies are replacing traditional methods because of their accuracy and reliability. This technological shift has enabled researchers and food industries to develop efficient methods for evaluating the quality, physical characteristics, internal damage, variety, and defects of seeds.

Table 2.1 presents the key methodologies for assessing physical quality in reviewed literature.

Table 2.1: Key methodologies for assessing physical quality in reviewed literature

Year	Methodology	Limitations/ Demerits
2012 [6]	- Counting sound-quality soybeans from a digital color image visualizing clustered seed arrangement. - The authors applied morphological opening and closing operations to disconnect physically touching seeds in the acquired images.	- Morphological opening and closing cause a significant loss of information about the boundary of seeds. - The operations may completely open small-size objects from the image. - The proposed method cannot be extended to a real-world sample of harvested seeds as the sample contains foreign objects as well.
2017 [9]	- Classification of soybeans from a digital color image visualizing clustered seed arrangement. - The authors employed various image processing methods for different purposes, such as the threshold method for segmenting the foreground and background and median blur to smooth images. - They applied methods like morphological opening, shape discrimination, image subtraction, boundary detection, and image filtering to detect different types of objects in harvested soybeans.	- The authors captured seed sample images from one view, missing information about seeds from the hidden view. - The morphological opening and closing cause a loss of information about the boundaries of objects. Additionally, the watershed algorithm may over-segment or under-segment images depending on changes in brightness, contrast, or color features. Similarly, image subtraction removes object boundaries, resulting in a loss of part of the object information. - Also, the precision and accuracy in threshold value selection for image processing depend on human expertise.
2018 [10]	- Classification of sound-quality white maize and thirteen materials undesired with maize using a	- The setup for multispectral imaging is more expensive and complex than simple RGB color imaging.

	multispectral imaging method with a wavelength range from 375 to 970 nm. - They made a specific arrangement with the Ulbricht sphere and camera to place maize samples. - LEDs illuminated the setup to mitigate object shadows and the discriminant analysis with the least squares method to analyze the objects.	- The accuracy of results is dependent on the controlled lighting conditions. The lighting variations affect spectral accuracy. - Processing features is computationally intensive and requires specialized analysis techniques. - Usually, multispectral imaging captures seed sample images from one view, missing information about seeds from the hidden view.
2018 [11]	- Study of the internal structure of harvested wheat of Rosa, Kama, and Canadian varieties. - Wheat grains manually arranged in disconnected positions for X-ray imaging and scanned the X-ray photograms using a flatbed scanner. - The physical characteristics of the wheat grains were analyzed using PCA and multivariate factor analysis methods.	- X-ray imaging is designed to extract information about the internal structure of objects. - It provides grayscale images and hence does not extract much information about features such as color and defects of the physical surface view. - It may not distinguish between surface defects or texture and the internal structure of seeds. - The X-ray imaging setup is costly and requires radiation safety protocols with trained operators.
2022 [13]	- An image segmentation and object classification methodology for the quality assessment of soybean seeds. - The authors created a ground-truth image dataset for image segmentation using Mask R-CNN, while for classification, they designed their lightweight neural network, Soybean Network.	- The authors did not consider contamination of foreign objects in the seed samples. - The Mask R-CNN requires a large amount of annotated data. Preparing such a dataset is a tedious task. - Also, Mask R-CNN requires excessive memory in the training and inference phases, as it is computationally intensive. - Mask R-CNN does not perform well on the classification of unseen object categories like unseen foreign objects.
2023 [14]	- The authors presented an approach for the fast classification of soybean seeds using the MSRCR for segmenting images captured under uneven lighting conditions.	- MSRCR is a hand-crafted algorithm, not adaptive. It may suffer from visual artifacts. - Also, MSRCR exhibits inconsistent performance with varying illumination conditions.

	- They optimized a lightweight convolutional neural network for quick classification of soybean seeds into four classes based on quality, abnormality, damage, etc.	- The multi-scale method needs multiple Gaussian convolutions, making MSRCR computationally expensive. - A GPU is required to implement MSRCR in real-time applications. - The authors considered only four classes of soybeans.
2024 [15]	- The hyperspectral imaging in the wavelength range of 930 to 1866 nm was implemented to detect insect-infected maize seeds. - The authors applied the GLCM to extract the texture features from the ratio of 1160 to 1310 nm as it exhibited maximum variance. - They fed the extracted features to a hybrid neural network of a 1-dimensional CNN and a bidirectional Long Short-Term Memory model.	- The setup for multispectral imaging is more expensive and complex than simple RGB color imaging. - The lighting variations affect spectral accuracy. - Images captured from one view, missing information about seeds from the hidden view. - The proposed method focuses only on identifying the insect infested seeds. The performance of the proposed method in the presence of other anomalies is not tested.

2.1 Problem Definition

The present work focuses on the evaluation of the quality of dry beans/ seeds using image processing and artificial intelligence.

2.2 Objectives

The objectives of the present work are as follows.

- Detection of objects present in a seed sample using morphological image processing techniques with an accuracy above 95%.
- Recognition and classification of detected objects present in a sample as per ‘Indian Grading Standard’ with an accuracy above 90%.

Chapter 3

LINEAR PATH INTERPOLATION FOR OBJECT DETECTION

3.1 Computational Flow for the Proposed Image Processing Scheme

Figure 3.1 illustrates the computational flow of the proposed image processing-based automatic system to evaluate the physical characteristics of harvested beans and seeds [6-9].

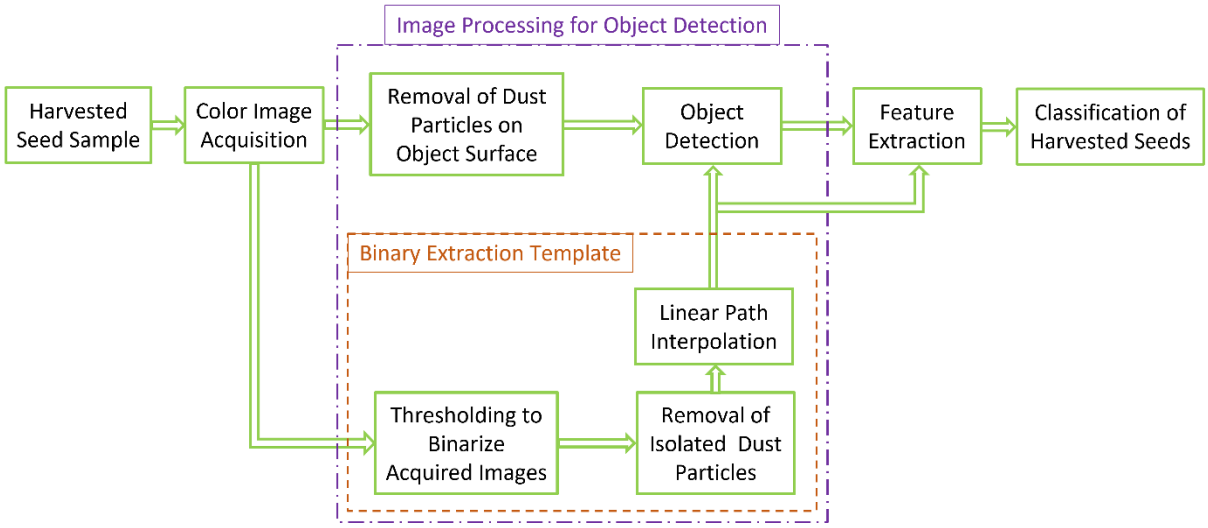

Figure 3.1: Computational flow for the proposed image processing scheme

3.2.1 Sampling Harvested Soybeans

For experimental purposes, a random sample of harvested soybean seeds weighing 700 grams was obtained from local farmers in Shirpur, Maharashtra (Pin 425405). It was harvested in January 2021 and had a moisture content of approximately 12%. Being a random sample, it included sound-quality seeds, seeds with potential defects, and foreign contaminants.

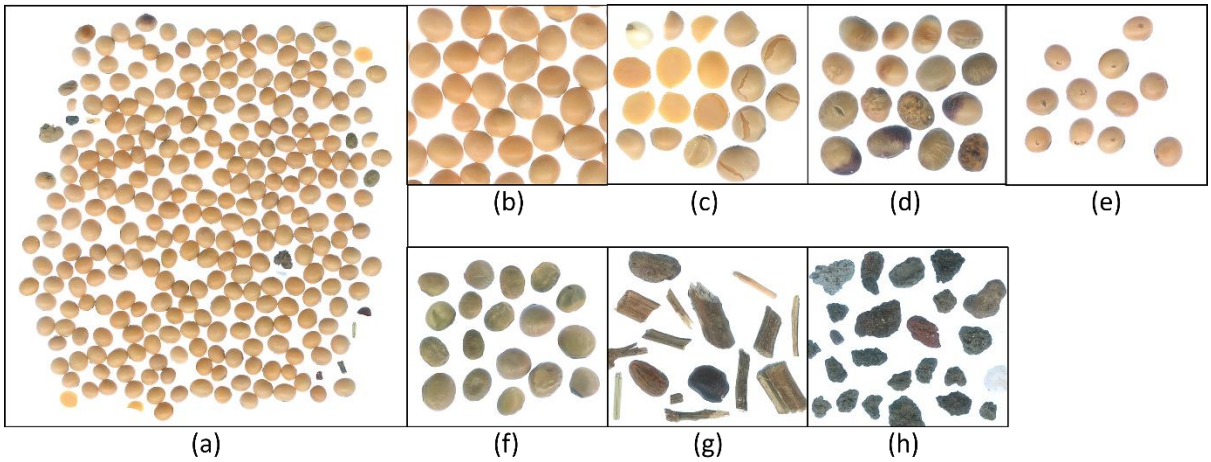

Figure 3.2: (a) Sample of harvested soybean seeds, (b) Sound-quality soybean seeds, (c) Cracked or split or broken soybean seeds, (d) Damaged or discolored soybean seeds, (e) Insect-infected soybean seeds, (f) Immature or shrivelled soybean seeds, (g) Organic foreign materials, (h) Inorganic foreign materials

The defective seeds were categorized by their physical characteristics under the guidance of agricultural experts. The seeds were manually inspected and sorted into eight groups based on their quality and defects, including sound-quality seeds, cracked seeds, split seeds, broken seeds, damaged or discolored or insect-infected seeds, immature or shriveled seeds, organic foreign materials, and inorganic foreign materials. Refer to Figure 3.2 for an illustration of the physical appearance of soybean seeds for each category.

3.2.2 Two-Sided Color Image Acquisition

Capturing images of harvested seed samples is a crucial step in the automatic quality evaluation process, as it is responsible for acquiring maximum visual information about every seed in the sample. In the present work, a flatbed scanner with the model number HP Scanjet G4050 was utilized to scan clear and noise-free images of harvested seeds. Figure 3.3 shows the experimental setup for image acquisition. It consists of an ultra-clear, 9H hardness Polyethylene Terephthalate glass sheet to hold the seed sample. A plain white background facilitated better contrast between the seeds, their anomalies, and the background. The backlight system, with no noticeable flicker, mitigated shadows of seed samples during scanning. The complete setup was operated in an enclosed environment to capture seed sample images under uniform conditions of distance, lighting, and background.

To manage the available size of the imaging platform of the HP Scanjet G4050 scanner, the eight groups of harvested soybeans, already sorted in the previous stage, were divided into smaller chunks, each chunk having an approximate weight of 20 to 40 grams. These chunks were randomly placed in a clustered form on the imaging platform without any specific arrangement. The clustered form saves time for image acquisition and the number of process iterations. The sample chunks were scanned from the bottom and top views to capture two-sided images of each seed. The flatbed scanner was positioned in its regular position, i.e., below the seed sample, to capture the bottom view of the seeds. Similarly, the flatbed scanner was inverted, i.e., positioned upside down, to capture the top view of the seeds. Figure 3.4 shows the bottom view and top view images acquired for a soybean seed sample. This approach captured information about the physical characteristics on both sides of the sample seeds without disturbing the seeds' position or causing any movement. Moreover, acquiring images by placing the seed samples in a specific marking area on the imaging platform maintained dimensional uniformity for the two views of the seeds. It also addressed the problem of coordinate matching between the two corresponding views of the seeds from both images. The flatbed scanner settings were configured to scan seed samples for 2D images of size 3500×3500 pixels at a resolution of 300 pixels per inch in color format.

As discussed earlier, while implementing the proposed two-sided image acquisition system, we used the HP Scanjet G4050 flatbed scanner. However, this scanner was not originally designed for scanning images in an inverted (upside-down) position. Due to this, the scanning head can sometimes become stuck along its roller path when used in an inverted position a few times. We resolved this issue by opening the scanner and fixing the problem from time to time.

This adjustment was made as there was no proper method available to capture two-sided images of harvested seed samples with good resolution, picture quality, and minimum noise.

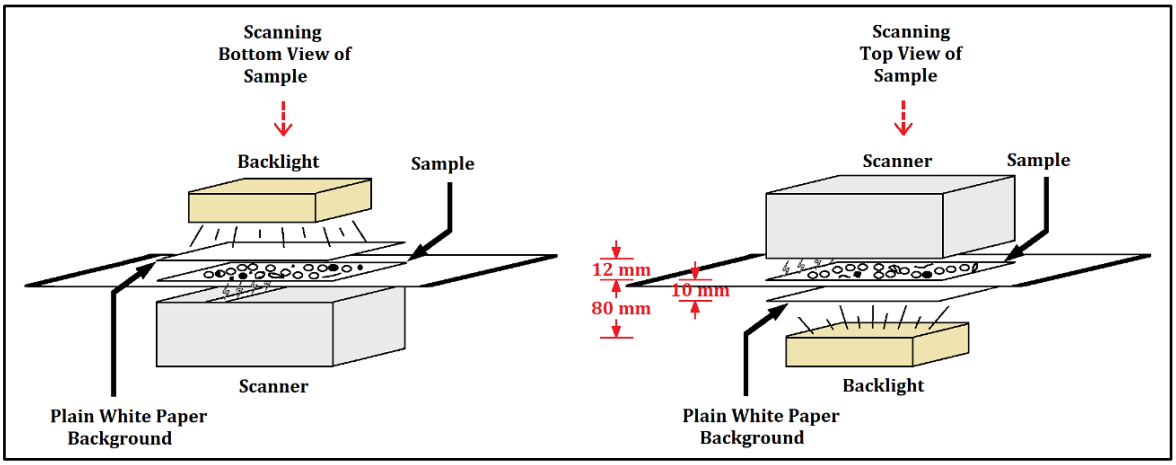

Figure 3.3: Schematic for two-sided photo-scanning of harvested seed sample

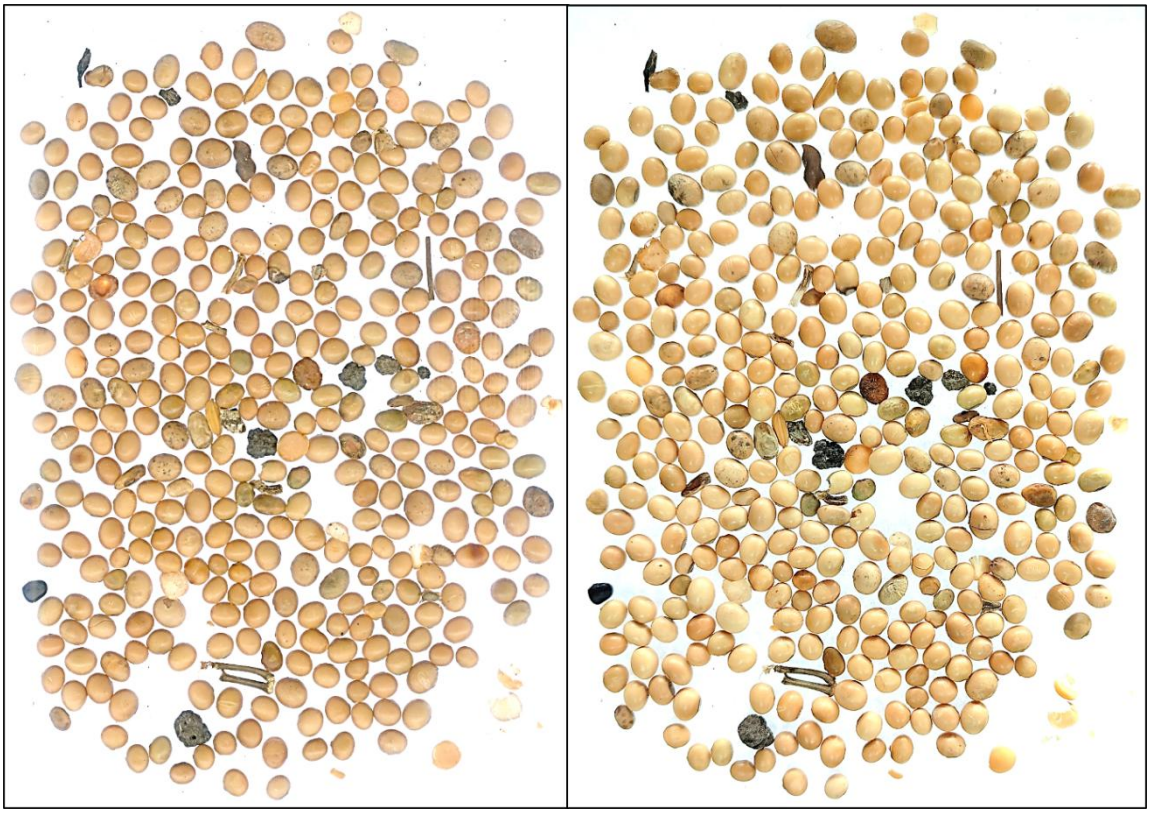

Figure 3.4: The acquired (a) Bottom view and (b) Top view images of a soybean seed sample

3.2.3 Thresholding to Binarize Acquired Images

The white background contrasted well with the sample seeds. Therefore, it was possible to use the thresholding method to segment harvested seeds from the background. For this purpose, the RGB color image of either the top view or bottom view was converted to the HSV format,

and a single threshold value, also known as a global threshold, was applied to the V-channel of the HSV image. Global thresholding is the image processing-based simplest method, significantly reducing computations in the image segmentation process. This operation produced a binary image that represented harvested seeds in white pixels and the remaining area in black pixels. Equation (1) expresses the mathematical equation for the global thresholding method.

$$B(i,j) = \begin{cases} 1, & If\ V(i,j) < T_v \\ 0, & Otherwise \end{cases} \quad (1)$$

Where,

B(i,j) represents a pixel with coordinates *(i,j)* in the binary image,

V(i,j) represents a pixel with coordinates *(i,j)* in V-channel of the image in HSV format and

T_v represents the global threshold value

3.2.4 Removal of Dust Particles

In the present work, the authors observed that typically, there are no adverse noises appearing in the acquired images due to the controlled environment of the image acquisition setup. However, the dust particles present in harvested seed samples manifest as noise in the captured images. To address the issue, the authors processed the acquired images using two noise removal methods, one based on thresholding to remove dust particles utilizing the binary form of the acquired images and the other based on median filtering utilizing the color format of the acquired images [6-12].

In the thresholding method, a 25×25 pixels sliding window was moved across the binary image obtained in the previous step pixel by pixel, capturing objects fitting within the window. These objects were identified as dust particles and removed by changing their (white) pixels to black, making those objects the background. So that they are not detected and processed during the subsequent object detection processes. The method specifically targeted independent dust particles in the images. Figure 3.5 illustrates the outcome of the thresholding method.

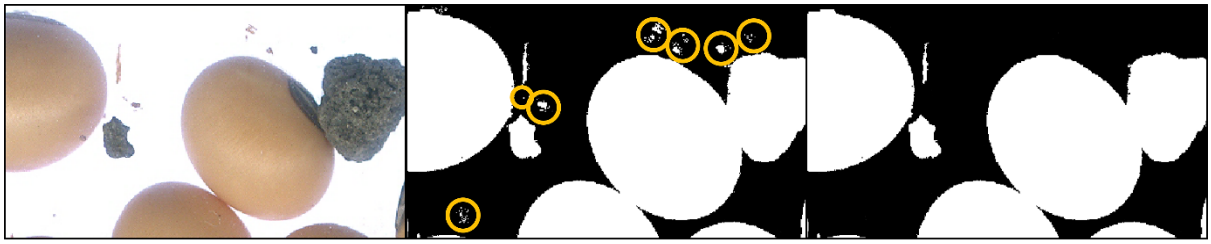

Figure 3.5: (a**)** Acquired image of soybean seed sample with dust particles, (b**)** Binary form of the image (orange circles mark the dust particles), (c) Binary form of the image with dust particles removed

The median filtering method focused on dust particles on the surfaces of the objects. In this method, a 5×5 pixels sliding window was applied to the H, S, and V channels of the image in

HSV format pixel by pixel. The algorithm computed the median of all pixel intensity values in this window. For each computation, the image pixel corresponding to the center pixel of the window was replaced with the median value. Figure 3.6 illustrates the outcome of the median filtering.

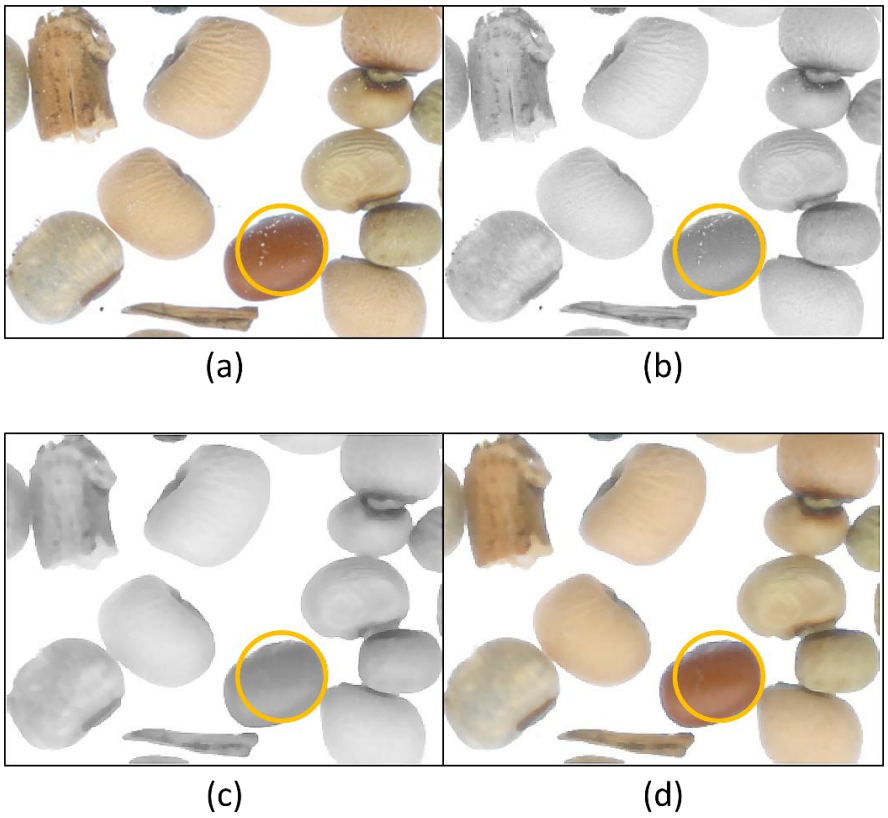

Figure 3.6: (a**)** Acquired image of the black-eyed peas sample with dust particles, (b**)** V-channel of the image, (c) V-channel after median filtering, (d) Output color image after median filtering (orange circles indicate the presence of dust particles before filtering and their removal afterward)

3.2.5 Linear Path Interpolation (LPI)

The random placement of harvested seed samples in a clustered form on the imaging platform results in physical contact between neighboring seeds of the sample. Therefore, the sample seeds in the acquired images appear as if physically touching their neighboring seeds. To evaluate the physical quality of the entire seed sample, it is necessary to assess the physical characteristics of every seed in the sample. Therefore, it is crucial to detect every individual seed, seed anomaly, or foreign material in the seed sample images.

In the proposed work, we developed a novel image processing-based method called Linear Path Interpolation (LPI) to minimize information loss during the object detection process and overcome the limitations of existing methods. The proposed method extracts a binary mask for each object in the acquired image. However, the effectiveness of the binary mask depends on its extraction process. In the proposed method, we utilize the topology of physically touching

objects to extract binary masks, as we observed that it is a robust and good feature for this purpose.

Topology of Physically Touching Objects

As illustrated in Figure 3.7, the physically touching objects from a harvested seed sample constitute V-grooves at the region of their physical touch. It is ideal to separate such physically touching objects over the path of physical touch only. Therefore, accurately identifying the region where these objects touch is crucial. However, tracking the exact path of physical touch between the objects is difficult. This is because of the similar intensity values of pixels as the region of physical touch. Therefore, instead of tracking the path of physical touch, the present work follows an alternative approach of rebuilding it. For this purpose, in the proposed work, the authors observed that the region of physical touch between the harvested seeds typically follows a straight line ending at the corners of the V-grooves. For clarity in the discussion, we refer to these pixels as X_A and X_B. This topology of the physically touching objects plays a pivotal role in designing the proposed method. By identifying valid pairs of pixels X_A and X_B, it is possible to recreate the linear paths representing the regions of physical touch.

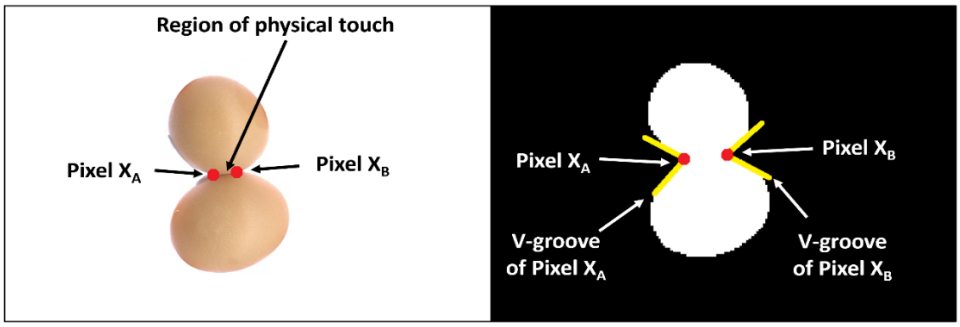

Figure 3.7: Soybean seeds with (a) Region of physical touch between pixels X_A and X_B, (b) V-grooves of pixels X_A and X_B

This operation involves the following steps.

Step 1: Finding V-grooves of the regions of physical touch

Step 2: Selecting legitimate combinations of pixels X_A and X_B

Step 3: Rebuilding region of the physical touch

The following subsections provide a detailed discussion of these steps.

Step 1: Finding V-grooves of the regions of physical touch

V-grooves created by physically touching objects can have any direction of orientation. Moreover, due to randomness in the shape and size of the seed sample objects, V-grooves can have any angle and length of their edges. These V-grooves may align with one of the eight cardinal or intercardinal directions, North, East, South, West, North-East, South-East, South-West, and North-West, or they may not fall on a standard alignment. Figure 3.8 illustrates a few of the random V-grooves created by the physically touching objects.

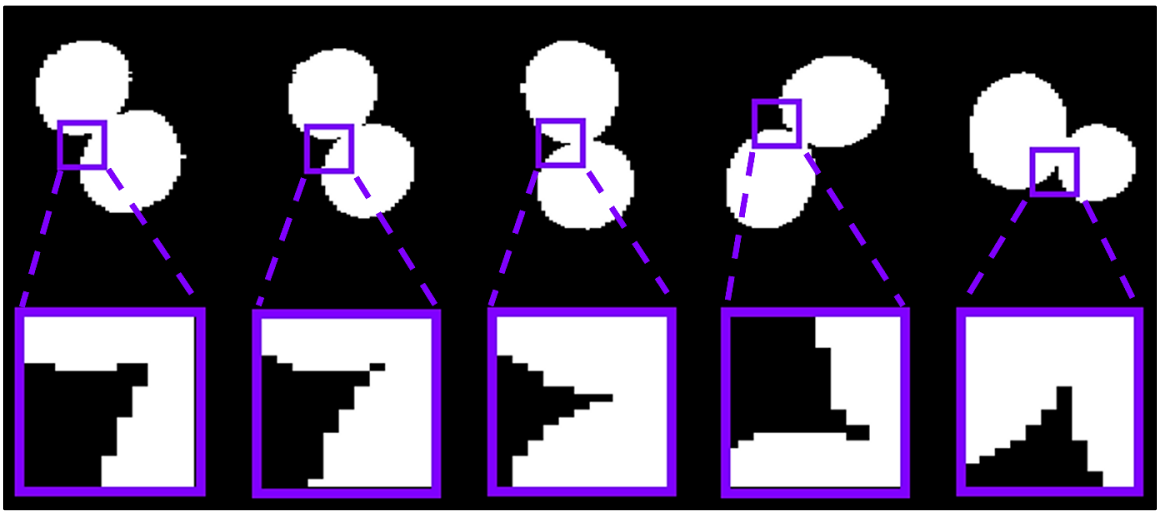

Figure 3.8: Examples of V-grooves of pixels X_A or X_B

To address this problem, the present work adopted the template-matching method. For this purpose, eight templates, each designated for detecting V-grooves aligned with one of the eight cardinal or intercardinal directions of the world, were designed. Figure 3.9 illustrates the association of eight V-grooves considered in this work. V-grooves not aligned perfectly with any of the eight directions were analyzed through computations to find their closest matching V-groove from the eight directions. Based on this analysis, non-aligned V-grooves were considered for a direction from the eight directions for further processing.

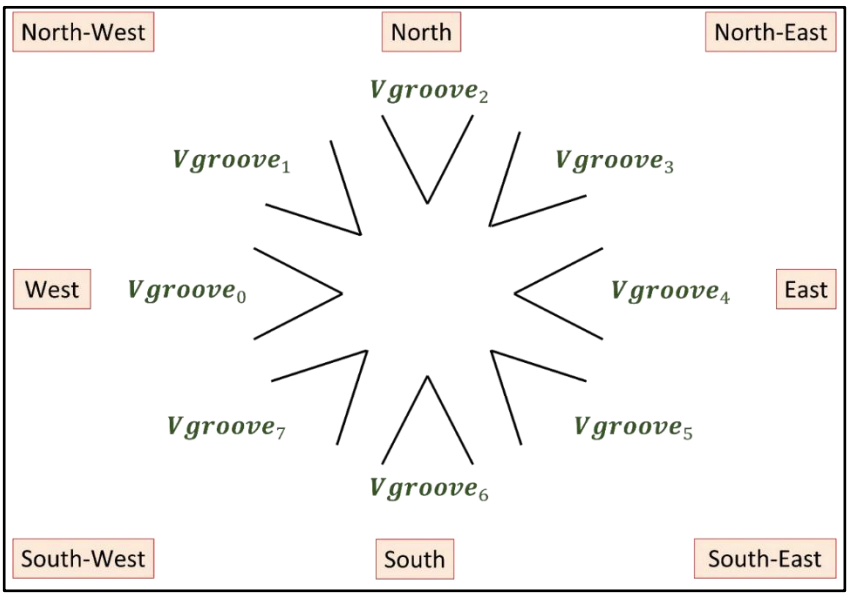

Figure 3.9: V-grooves in eight directions

Figure 3.10 illustrates the templates designed in the proposed method to identify all V-grooves in seed sample images that are aligned or not aligned with any of the eight directions. For example, the template named S_0 is designed to find V-grooves that match the size, shape, angle, and other parameters of Vgroove0 in the West direction. However, these parameters do not follow a hard limit. Hence, the template S_0 is capable of finding V-grooves aligned between $Vgroove_0$ and its neighbor $Vgroove_1$ and between $Vgroove_0$ and its neighbor $Vgroove_7$. To

clarify this confusion, the V-groove found by S_0 and aligned between $Vgroove_0$ and $Vgroove_1$ is considered for $Vgroove_0$ or $Vgroove_1$, depending on the highest level of matching. Similarly, the V-groove found by S_0 and aligned between $Vgroove_0$ and $Vgroove_7$ is considered for $Vgroove_0$ or $Vgroove_7$, depending on the highest level of matching. This same process applies to the other templates, such as S_1, S_2, S_3, S_4, S_5, S_6, and S_7, as illustrated in Figure 3.10. These templates operate on a binary image generated from either the top or bottom view of the seed samples, reducing the computation cost of processing the color images otherwise.

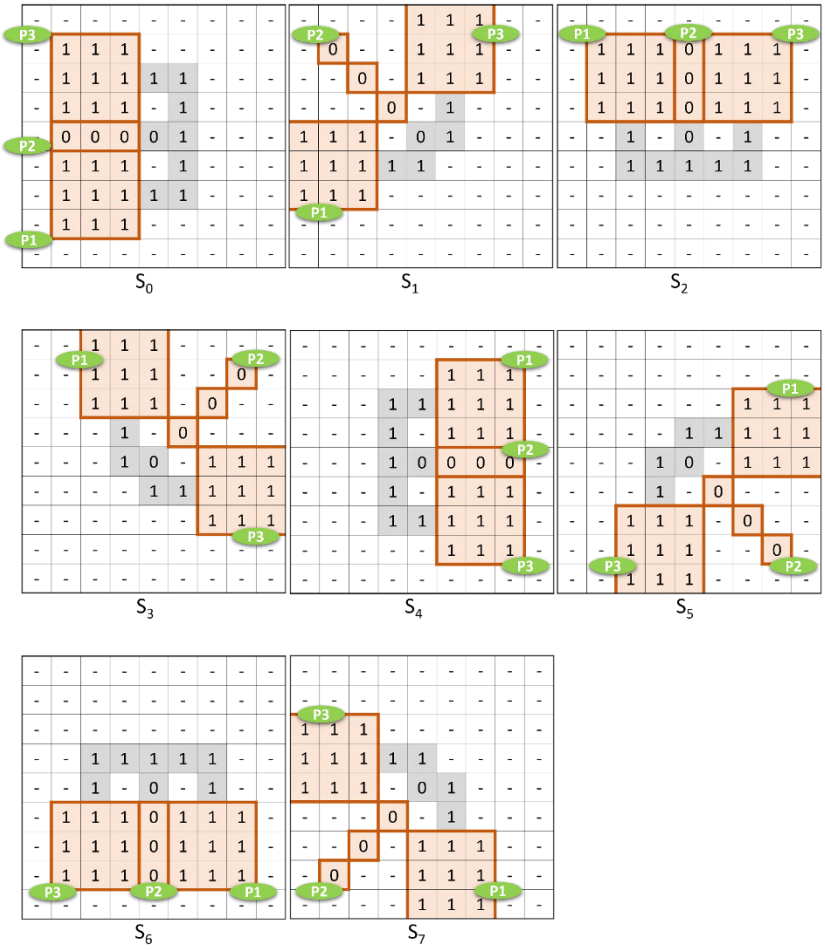

Figure 3.10: Eight templates (pixel '-' can take any binary bit value)

This decision-making depends on the level of matching between the V-grooves found in the binary image and the reference V-grooves from the eight cardinal and intercardinal directions. To compute the level of matching, the templates S_0, S_1, S_2, S_3, S_4, S_5, S_6, and S_7 are designed with specific bit patterns. As illustrated in Figure 3.10, the 9×9 pixels templates consist of orange patterns P1, P2, P3, and a gray pattern. Pixel '-' can take any binary bit value. Each template moves through the binary image from top to bottom and pixel by pixel to find the V-grooves. While moving through the binary image, the template uses its gray pattern to perform initial computations and check whether the visited part in the binary image contains a V-groove. In this process, if the gray pattern matches a pattern in the visited part of the binary image, then the visited part is considered a potential V-groove. Then, it computes the counts of 1s and 0s for the visited part of the image in correspondence with the orange patterns P1,

P2, and P3 to check whether the counts satisfy the conditions in Table 3.1. Based on the conditions satisfied by the computed counts, decisions are made about categorizing the detected Vgroove by following the rules outlined in Table 3.1. After categorizing the detected V-groove for a specific orientation, the pixel at the corner of this V-groove, which also corresponds to pixel 0 of the gray pattern when matched, is stored as a black pixel in a blank white pixels' image designated for storing the corner pixels of V-grooves detected for that specific category.

Table 3.1: Conditions and rules to validate and decide orientation for V-grooves detected using j^{th} template

Condition	Rule for decision-making
Condition 1: $\left[Count\left(P_2^{(S_j)}\right) > 0\right] \wedge \left\{\left[Count\left(P_1^{(S_j)}\right) \geq 4\right] \vee \left[Count\left(P_3^{(S_j)}\right) \geq 4\right]\right\} \wedge \left\{\left\lvert Count\left(P_1^{(S_j)}\right) - Count\left(P_3^{(S_j)}\right)\right\rvert \leq 3\right\}$	The pixel in the visited part of the binary image corresponding to 0 in the visiting template S_j is a pixel at the corner of the detected V-groove of orientation $Vgroove_{(j)}$.
Condition 2: $\left[Count\left(P_2^{(S_j)}\right) > 0\right] \wedge \left\{\left[Count\left(P_1^{(S_j)}\right) \geq 4\right] \vee \left[Count\left(P_3^{(S_j)}\right) \geq 4\right]\right\} \wedge \left\{\left\lvert Count\left(P_1^{(S_j)}\right) - Count\left(P_3^{(S_j)}\right)\right\rvert > 3\right\} \wedge \left\{Count\left(P_1^{(S_j)}\right) < Count\left(P_3^{(S_j)}\right)\right\}$	The pixel in the visited part of the binary image corresponding to 0 in the visiting template S_j is a pixel at the corner of the detected V-groove of orientation $Vgroove_{(j-1,mod8)}$.
Condition 3: $\left[Count\left(P_2^{(S_j)}\right) > 0\right] \wedge \left\{\left[Count\left(P_1^{(S_j)}\right) \geq 4\right] \vee \left[Count\left(P_3^{(S_j)}\right) \geq 4\right]\right\} \wedge \left\{\left\lvert Count\left(P_1^{(S_j)}\right) - Count\left(P_3^{(S_j)}\right)\right\rvert > 3\right\} \wedge \left\{Count\left(P_1^{(S_j)}\right) > Count\left(P_3^{(S_j)}\right)\right\}$	The pixel in the visited part of the binary image corresponding to 0 in the visiting template S_j is a pixel at the corner of the detected V-groove of orientation $Vgroove_{(j+1,mod8)}$.
Note: If none of the above conditions is satisfied, the pixel in the visited part of the binary image corresponding to 0 in the visiting template S_j is not considered a pixel at the corner of a V-groove, and hence, it is discarded.	

As mentioned in Table 3.1, template S_j moving through the binary image from top to bottom and pixel by pixel identifies which reference orientation the detected V-groove belongs to. Whether it belongs to the reference orientation $Vgroove_{(j-1, mod8)}$, $Vgroove_j$, or $Vgroove_{(j+1, mod8)}$, where $0 \leq j < 8$. For Example, a binary image contains eleven seeds, ten seeds physically touching another seed, forming ten V-grooves at the region of physical touch, as illustrated in Figure 3.11. Template S_j, given j=7 (that is S_7), slides from top to bottom and pixel by pixel through this binary image, searching for the V-grooves $Vgroove_6$, $Vgroove_7$, and $Vgroove_0$. Based on the satisfaction of condition 1 in Table 3.1, template S_7 successfully searches V-

grooves, as shown in Figure 3.11(a) and Figure 3.11(b). Similarly, template S_0 slides from top to bottom and pixel by pixel through this binary image, searching for the V-grooves $Vgroove_7$, $Vgroove_0$, and $Vgroove_1$. Based on the satisfaction of condition 1 in Table 3.1, template S_0 successfully searches a V-groove, as shown in Figure 3.11(c). Likewise, template S_1 finds a V-groove, as shown in Figure 3.11(d). Template S_6 successfully finds a V-groove, as shown in Figure 3.11(e). However, there are two adjacent black pixels forming the corner of the V-groove in Figure 3.11(e). In such a case, template S_6 finds both as the potential pixels at the corner of the V-groove to be used for subsequent operation of rebuilding the region of physical touch.

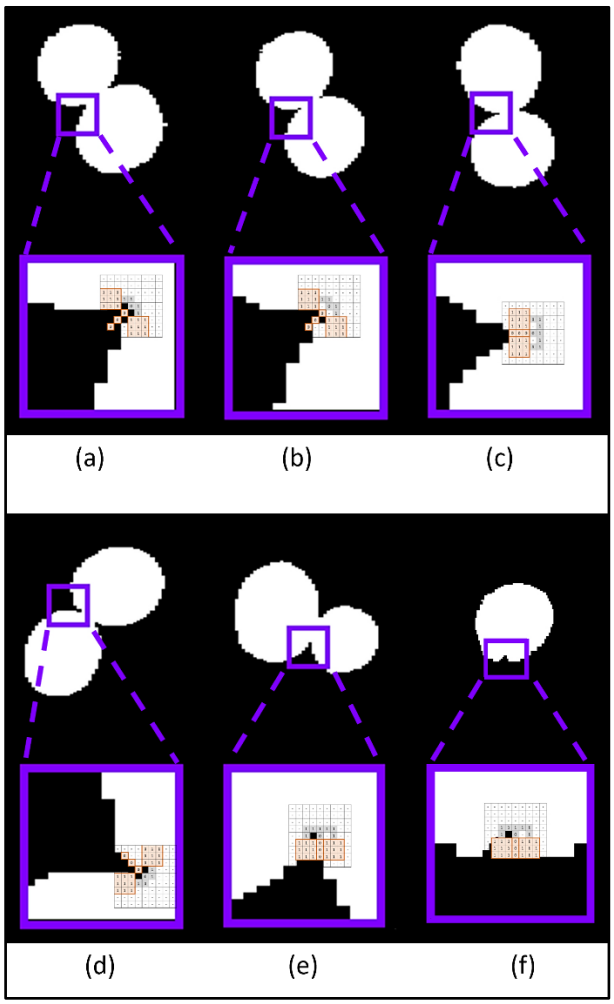

Figure 3.11: Detection of V-grooves using pattern windows

In Figure 3.11(f), the gray pattern of template S6 matches the visited part in the image. However, it fails to satisfy the conditions on counts of orange patterns P1, P2, and P3. Therefore, it is not identified as a valid V-groove.

Then, all the valid pixels identified at the corner of V-grooves are stored as black pixels in their designated blank white pixel images for further processing. The valid pixels identified at the corner of $Vgroove_0$ are stored as black pixels at their original coordinates in a blank white image called PV_0, which is designated to store corner pixels for $Vgroove_0$. The valid pixels identified at the corner of $Vgroove_1$ are stored as black pixels at their original coordinates in a

blank white image called PV_1, which is designated to store corner pixels for $Vgroove_1$. Likewise, the corner pixels of $Vgroove_2$, $Vgroove_3$, $Vgroove_4$, $Vgroove_5$, $Vgroove_6$, and $Vgroove_7$ are stored as black pixels at their original coordinates in their designated blank white images called PV_2, PV_3, PV_4, PV_5, PV_6, and PV_7 respectively.

Step 2: Selecting legitimate combinations of pixels X_A and X_B

The orientations of the V-grooves in a legitimate combination of pixels X_A and X_B always point in contrasting directions relative to each other. Additionally, these pixels maintain minimal spatial separation. The present work exploits this characteristic to recognize legitimate combinations of pixels X_A and X_B.

In this regard, the proposed algorithm visits every black pixel stored in images PV_0, PV_1, PV_2, PV_3, PV_4, PV_5, PV_6, and PV_7 one after another from top to bottom and pixel by pixel. Every pixel in PV_m (where $0 \leq m < 8$) being visited is momentarily labeled as X_A. During the visit, for a given X_A, the algorithm searches for its legitimate combination X_B in images PV_n, where $(m+3, mod8) \leq n \leq (m+5, mod8)$, as illustrated in Figure 3.12. For example, if X_A is being visited in PV_m (where $m = 1$), the algorithm will search for its legitimate combination X_B in images PV_n (where $n = 4, 5, 6$). Similarly, if X_A is being visited in PV_m (where $m = 4$), the algorithm will search for its legitimate combination X_B in images PV_n (where $n = 7, 0, 1$). Moreover, as per the characteristic of minimal spatial separation between a legitimate combination of pixels X_A and X_B, it is expected that X_B will be searched in PV_n for X_A in PV_m over a limited spatial distance from the X_A. Searching the entire image of PV_n is not required. For simplicity and clarity, we call the limited spatial distance to be searched the scanning area (SA). Our experimentation suggested that SA can be of size 75×75 pixels. Figure 3.12 shows the area SA_m for image PV_m containing X_A (where $0 \leq m < 8$). While searching for X_B, when a pixel demonstrates the shortest distance to X_A, it is provisionally designated as X_B. A provisional designation as X_B serves to verify against incorrect combinations of X_A and X_B. This is because the legitimacy of the provisional X_B cannot be determined until the algorithm has scanned all the PV_n images. Moreover, due to the random shapes of the objects, incorrect detection of V-grooves may have occurred in the previous step. This must be addressed in the current step. Otherwise, invalid combinations of X_A and X_B may be selected.

To ensure that the provisionally designated X_B is the legitimate X_B, a reverse verification procedure takes place. The reverse verification procedure searches for legitimate X_A of the provisionally designated X_B. Suppose that the provisionally designated X_B belongs to image $PV_n = PV_{(m+4, mod8)}$, then its legitimate X_A may belong to one of the images $PV_{(m+4+3, mod8)}$, $PV_{(m+4+4, mod8)}$, and $PV_{(m+4+5, mod8)}$. If the provisionally designated X_B belongs to image $PV_n = PV_{(m+5, mod8)}$, then its legitimate X_A may belong to one of the images $PV_{(m+5+3, mod8)}$, $PV_{(m+5+4, mod8)}$, and $PV_{(m+5+5, mod8)}$. In this reverse verification procedure, if the provisionally designated X_B finds the original X_A as its legitimate X_A, then only the given combination of X_A and X_B is selected as a legitimate combination. Otherwise, the provisionally designated X_B is ignored, and the algorithm continues its search to find the legitimate X_B for the X_A being visited. Furthermore, if no legitimate X_B is found for the X_A being visited, then the X_A is not considered

a legitimate X_A and is discontinued from further processing. The legitimate combinations of X_A and X_B are carried forward for further processing in the next step.

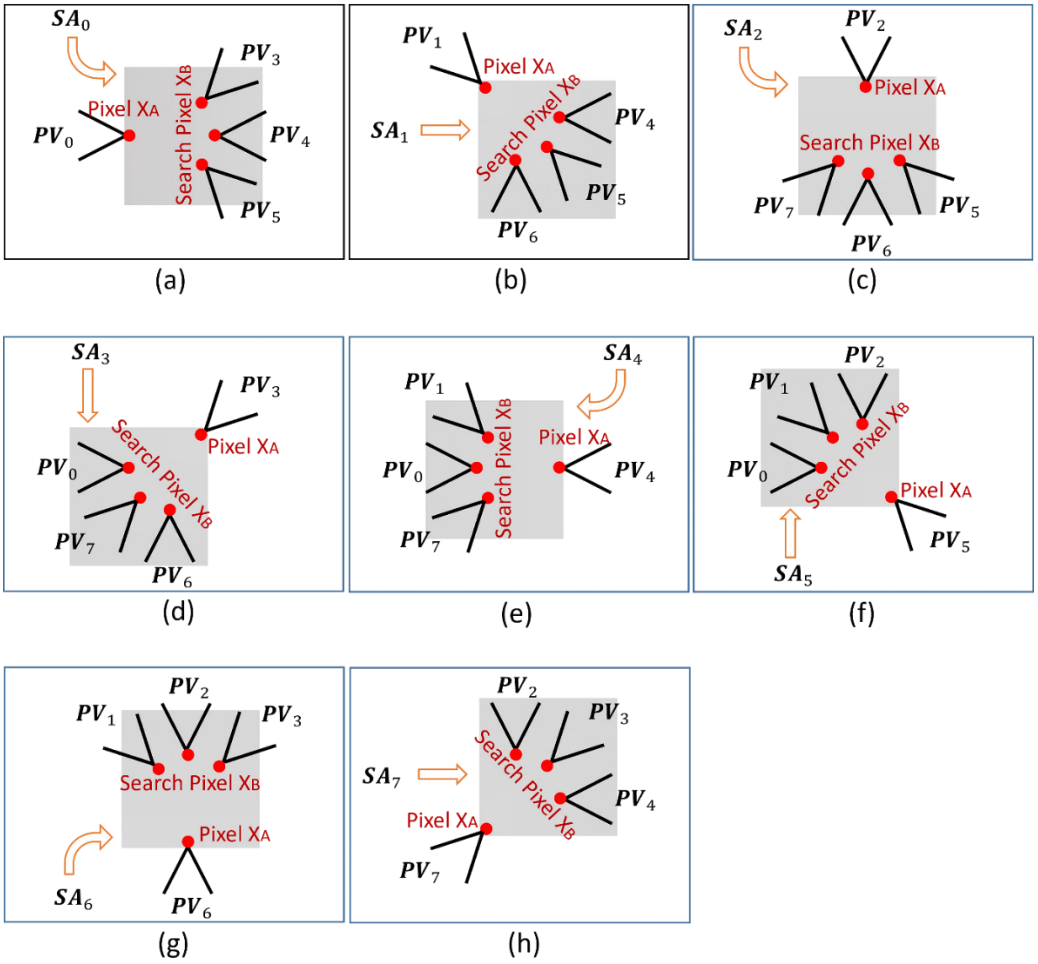

Figure 3.12: Search of X_B for X_A in (a) PV_0, (b) PV_1, (c) PV_2, (d) PV_3, (e) PV_4, (f) PV_5, (g) PV_6, (h) PV_7

Step 3: Rebuilding region of the physical touch

As shown in Figure 3.7, the region of physical touch between the harvested seeds is generally a straight line due to the elliptical shape of the harvested seeds. The proposed method utilizes this property to rebuild the path of physical touch. For this purpose, it builds a linear staircase between the X_A and X_B of a legitimate combination, as illustrated in Figure 3.13. To create the path, it is necessary to find the distance between X_A and X_B. Additionally, computations, such as determining the direction in which the path must proceed and the appropriate lengths for its horizontal and vertical lines, are essential.

For illustration, if (a_1, b_1) and (a_2, b_2) are the spatial coordinates of rows and columns for X_A and X_B, then the absolute distance value between X_A and X_B in terms of spatial coordinates is given by the mathematical expression in equations (2) and (3).

$$a_{mod} = |a_1 - a_2| \tag{2}$$

$$b_{mod} = |b_1 - b_2| \tag{3}$$

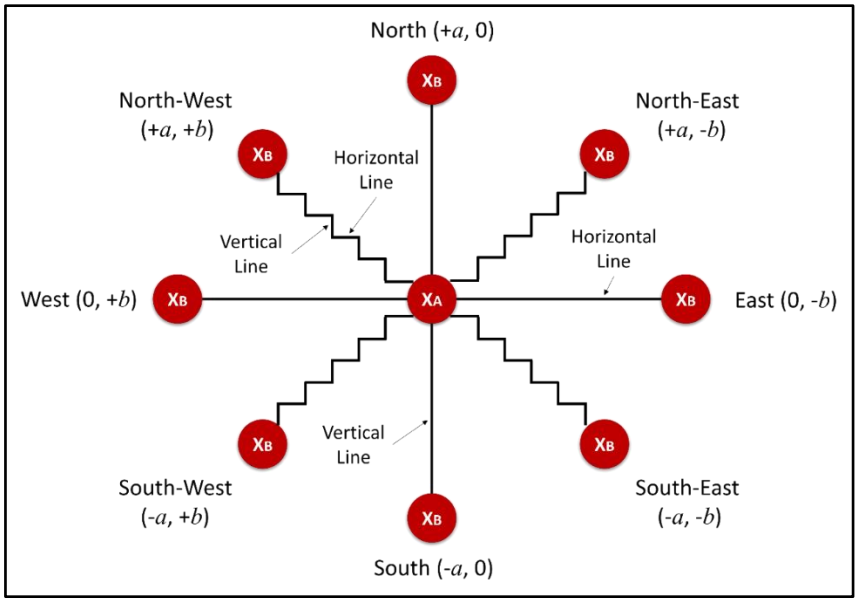

Figure 3.13: Scenarios for a linear path between pixels X_A and X_B

Table 3.2: Scenarios to fix the lengths of horizontal and vertical lines and selecting the starting line

Sr. no.	Scenario	Length of vertical line (in pixels)	Length of horizontal line (in pixels)	Start with
1.	$b_{mod} < a_{mod} \quad with \quad 0 < a_{mod}, b_{mod} < \infty$	$\left\|\frac{a_{mod}}{b_{mod}}\right\|$	1	Vertical line
2.	$a_{mod} < b_{mod} \quad with \quad 0 < a_{mod}, b_{mod} < \infty$	1	$\left\|\frac{b_{mod}}{a_{mod}}\right\|$	Horizontal line
3.	$a_{mod} = b_{mod} \quad with \quad 0 < a_{mod}, b_{mod} < \infty$	1	1	Vertical or horizontal line
4.	$a_{mod} = 0 \quad with \quad 0 < b_{mod} < \infty$	0	b_{mod}	Horizontal line
5.	$b_{mod} = 0 \quad with \quad 0 < a_{mod} < \infty$	a_{mod}	0	Vertical line

Based on the values of a_{mod} and b_{mod}, there are five scenarios for building the region of physical touch, as suggested in Table 3.2. These scenarios fix the lengths of horizontal and vertical lines and decide whether to start with a horizontal or vertical line.

Similarly, the non-absolute distance value between X_A and X_B in terms of spatial coordinates, as given by equations (4) and (5), provides a sign indicating the direction in which to proceed with building the path, as illustrated in Figure 3.13. There are eight cardinal or intercardinal directions.

$$a = a_1 - a_2 \quad (4)$$

$$b = b_1 - b_2 \quad (5)$$

3.2.6 Object Detection

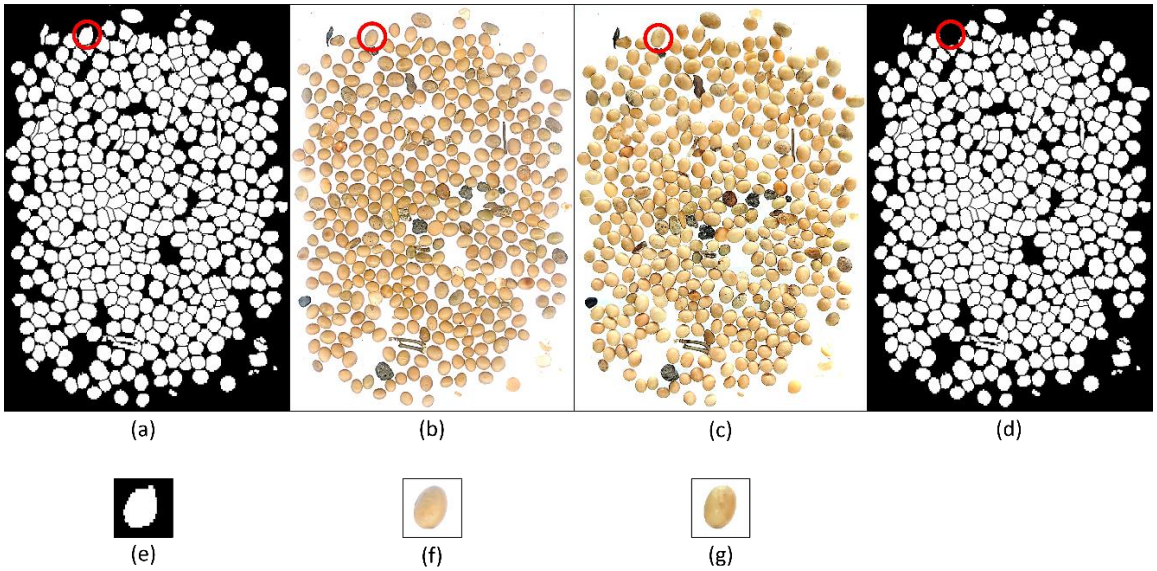

Figure 3.14: Images of a soybean seed sample showing (a) Binary mask, (b) Bottom view, (c) Top view, (d) Binary mask with deleted white area of the seed being processed, (e) White area of the seed being processed, (f) Bottom view of the seed being processed, (g) Top view of the seed being processed (red circle indicates the object being processed)

The binary image provided by the LIP algorithm in the last step consists of distinct objects from the seed sample. Figure 3.14(a) displays an example of such a binary image, showing separated objects for a soybean seed sample. The object detection block processes this image to retrieve each of these objects as an extraction template for finding its corresponding bottom view and top view from the acquired two-sided color images. Figure 3.14(e) displays an example of such an extraction template retrieved for an object highlighted in the red circle in Figure 3.14(a).

3.2.7 Feature Extraction and Classification of Harvested Seeds

The proposed method used the binary extraction template of every object like the one shown in Figure 3.14(e), its corresponding bottom view (Figure 3.14(f)), and its corresponding top view (Figure 3.14(g)). The extraction template provided geometric features, the bottom view and top view images provided color and texture details of the object.

The present work referred to the Indian Agmark grading standard for evaluating the physical characteristics of the harvested seeds. We considered harvested soybean seeds with eight classes. A two-level cascaded classification approach was employed. In the first level, the soybean seeds are classified into seven classes, including sound-quality/ cracked seeds, broken seeds, split seeds, damaged or discolored/ insect-infested seeds, immature or shriveled seeds, organic foreign materials, and inorganic foreign materials. The second stage focuses on further classifying the sound-quality/ cracked seeds into either sound-quality seeds or cracked seeds.

Chapter 4

HYBRID NEURAL NETWORK TO CLASSIFY HARVESTED BEANS

4.2.1 Sampling Harvested Black-Eyed Peas

For experimental purposes, a random sample of harvested black-eyed peas weighing 600 grams was obtained from local farmers in Dhule, Maharashtra. It was harvested in March 2024 and had a moisture content between 13% and 14%. It included good-quality peas, peas with visible physical defects, and foreign materials. The peas with visible defects were categorized by their physical characteristics under the guidance of professional agricultural experts. The sample was carefully sorted by hand into eight separate groups based on their quality and defects, including sound-quality peas, split peas, broken or cracked peas, damaged or discolored peas, insect-infected peas, immature or shriveled peas, organic foreign materials, and inorganic foreign materials. Refer to Figure 4.1 to visualize the physical appearance of harvested black-eyed peas and their anomalies. Table 4.1 presents a summary of the object counts and physical traits observed in the collected harvested black-eyed pea sample [5].

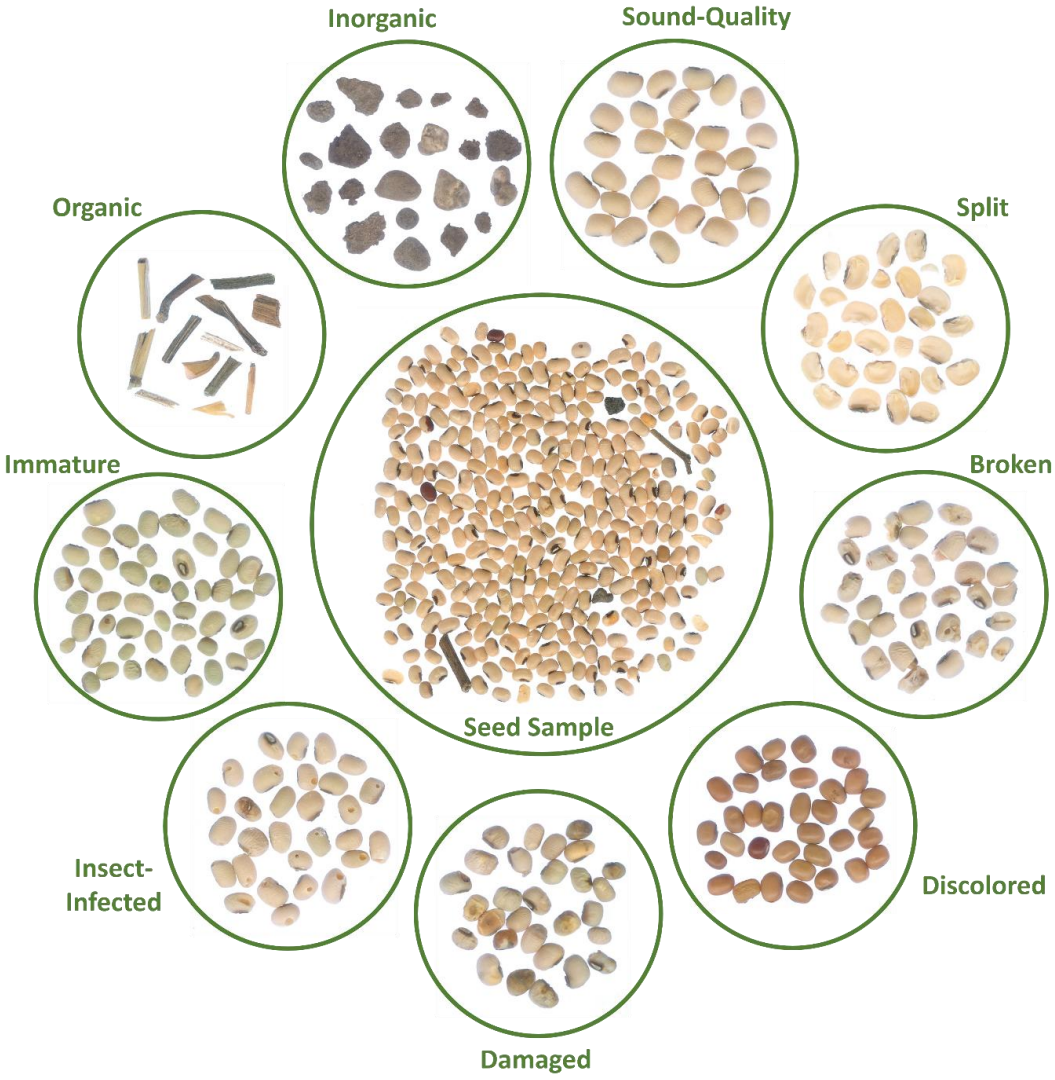

Figure 4.1: Harvested black-eyed peas with various anomalies and foreign materials

Table 4.1: Object count and physical traits in the collected harvested black-eyed pea sample

Category	Physical Traits	Count
Sound-quality	Completely developed black-eyed peas of curved oval shape and buff or straw-like color.	1849
Split	One of the parts after splitting the pea along its natural seam, no outer layer on the broken side, light buff or straw-like color, and of uneven shape.	486
Broken or cracked	Unbroken side is curved oval, the broken side is of uneven shape, with no outer layer on the broken side, light buff or straw-like color.	426
Damaged or discolored	Broken or unbroken kernels, showing signs of disease, discoloration, mold growth, or significant damage.	578
Insect-infected	Unbroken kernels without any damage, an insect-damaged area with holes.	377
Immature or shriveled	Broken or unbroken kernels of green-like color, not fully developed and may appear shriveled.	419
Organic foreign materials	Plant debris such as seed husks, stem fragments, straw pieces, and chaff, all in different colors and shapes.	532
Inorganic foreign materials	Soil clumps, sand particles, and small stones, all in different colors and shapes.	496

4.2.2 Color Image Acquisition

Figure 4.2 shows the setup we used to take pictures of the harvested black-eyed peas for this study. We used an HP Scanjet scanner, which has a CCD, connected to a computer for saving and processing the images. The scanner takes pictures at a resolution of 300 pixels per inch, resulting in 2D color images that are 3500 × 3500 pixels in size. We divided each batch of beans into smaller groups, with each group weighing no more than 40 grams so they would fit on the scanner. We placed plain white paper underneath to make the different beans stand out more clearly. To minimize shadows in the images, we used a set of 15-watt bulbs for backlighting, which didn't flicker. The distance from the bulbs to the background was adjusted between 7 cm and 11 cm to get varying light levels. We took pictures of each sample weighing up to 40 grams, making sure the beans were arranged on the scanning surface without overlapping. This way, we could capture all the details of the items.

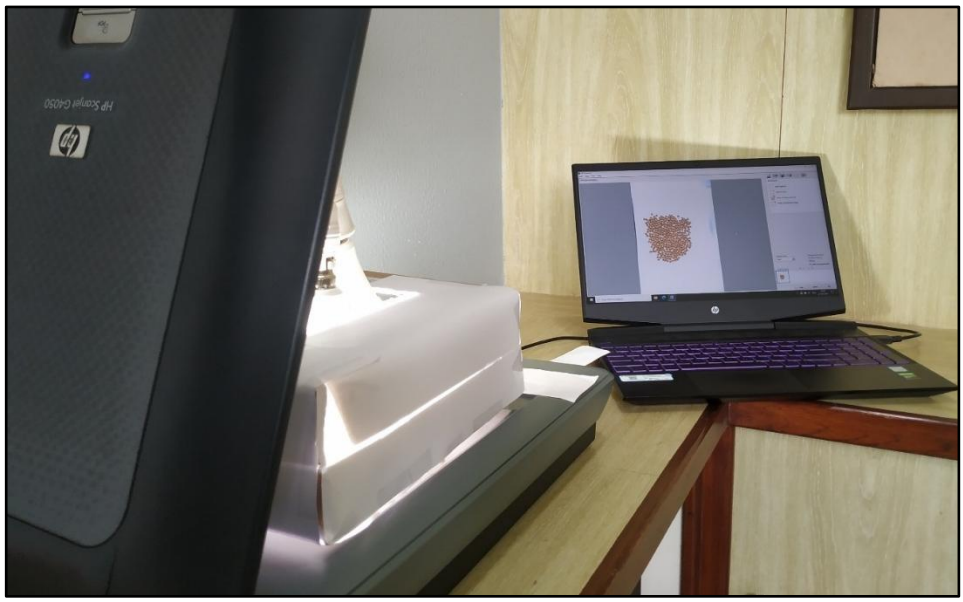

Figure 4.2: Color image acquisition of harvested black-eyed peas

4.2.3 Object Detection

Every image obtained from the scanner contained many objects. To implement artificial intelligence systems for classifying these objects, it was necessary to extract every object from the scanned images. To do this, the scanned images were converted to binary images using Otsu's thresholding method. After that, an image segmentation process was performed to detect and isolate individual objects. For this purpose, we employed the advanced SAM along with our proposed object retrieving module (see Figure 4.3) to accurately extract each object from the acquired images [16-18].

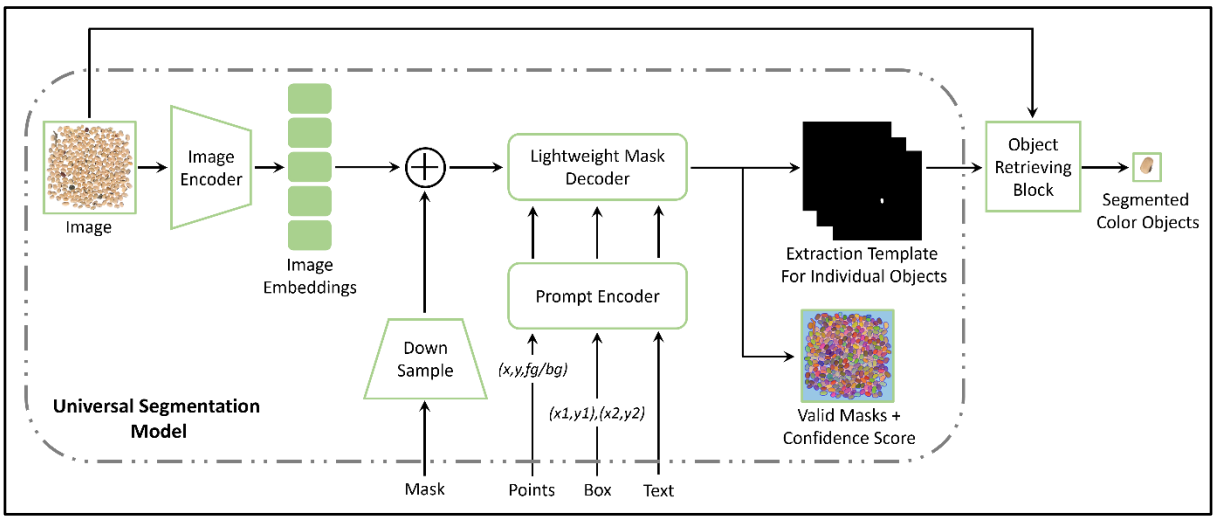

Figure 4.3: SAM equipped with custom object retrieving method [16]

4.2.4 Image Augmentation

Typically, harvested seed samples contain fewer defective seeds compared to those of good quality, resulting in an imbalanced dataset during training. This imbalance causes neural network models to overfit, favoring the dominant class and performing poorly when exposed

to new image datasets in the testing stage, even though it had shown good performance in the training stage. To tackle this issue, the present work implemented an image augmentation strategy using a generative approach called the Stable Diffusion. This method generated synthetic images specifically for underrepresented seed groups with fewer than 1000 instances.

Conventional augmentation techniques such as flipping, rotating, or cropping offer limited improvement in the diversity of training images. In contrast, we trained the stable diffusion model using RGB images of every class object to create a set of training samples maintaining the original diversity [19].

4.2.5 Dataset Split

After augmenting the acquired input images, every class had at least 1,000 images of individual object instances. A total of 1,000 images were randomly picked from each class to develop the classification model. In particular, for the class representing sound-quality soybeans, a subset of 1,000 images was chosen at random, as it had more than 1,000 images. Then, the chosen images were split into three separate datasets: 70% were allocated for training (i.e., 700 images), 20% for validation (i.e., 200 images), and the remaining 10% (i.e., 100 images) for testing.

4.2.6 Object Classification using EfficientNetV2B1+ViT Model

CNNs that are already trained on large datasets, such as ImageNet, have learned useful patterns and features from those datasets [20,21]. Commonly used pre-trained CNN architectures include MobileNetV2, ResNet50, VGG16, ViT, and EfficientNet, which are widely adopted for transfer learning tasks due to their proven performance and efficiency [6-15]. The performance of these models can be enhanced by implementing methods such as combining different models to form hybrid models based on transfer learning. It also requires merging features or applying attention techniques. These approaches help bring together the advantages of the various models involved in forming hybrid models. In our study, we developed an enhanced hybrid model that combines EfficientNetV2B1 [22,23] and ViT [24], utilizing CBAM [26] for the attention mechanism and feature merging.

Figure 4.4 illustrates the proposed enhanced hybrid architecture of EfficientNetV2B1 and ViT as implemented in this work. The architectural design comprises two parallel processing pathways. One for EfficientNetV2B1 and the other for ViT. Both pathways receive input images of size 224 × 224 pixels with three color channels for extracting features. The features from both processing pathways are merged through a fusion module and subsequently passed to the classifier block, which assigns the input image to one of eight target categories. The fusion layer applies a concatenation operation on the features received from the EfficientNetV2B1 and ViT networks [25]. To improve prediction accuracy, the proposed method employs an attention mechanism that helps the model focus on the most crucial parts of the input. In this study, the CBAM was implemented to help the model focus on the key features needed to distinguish harvested seeds.

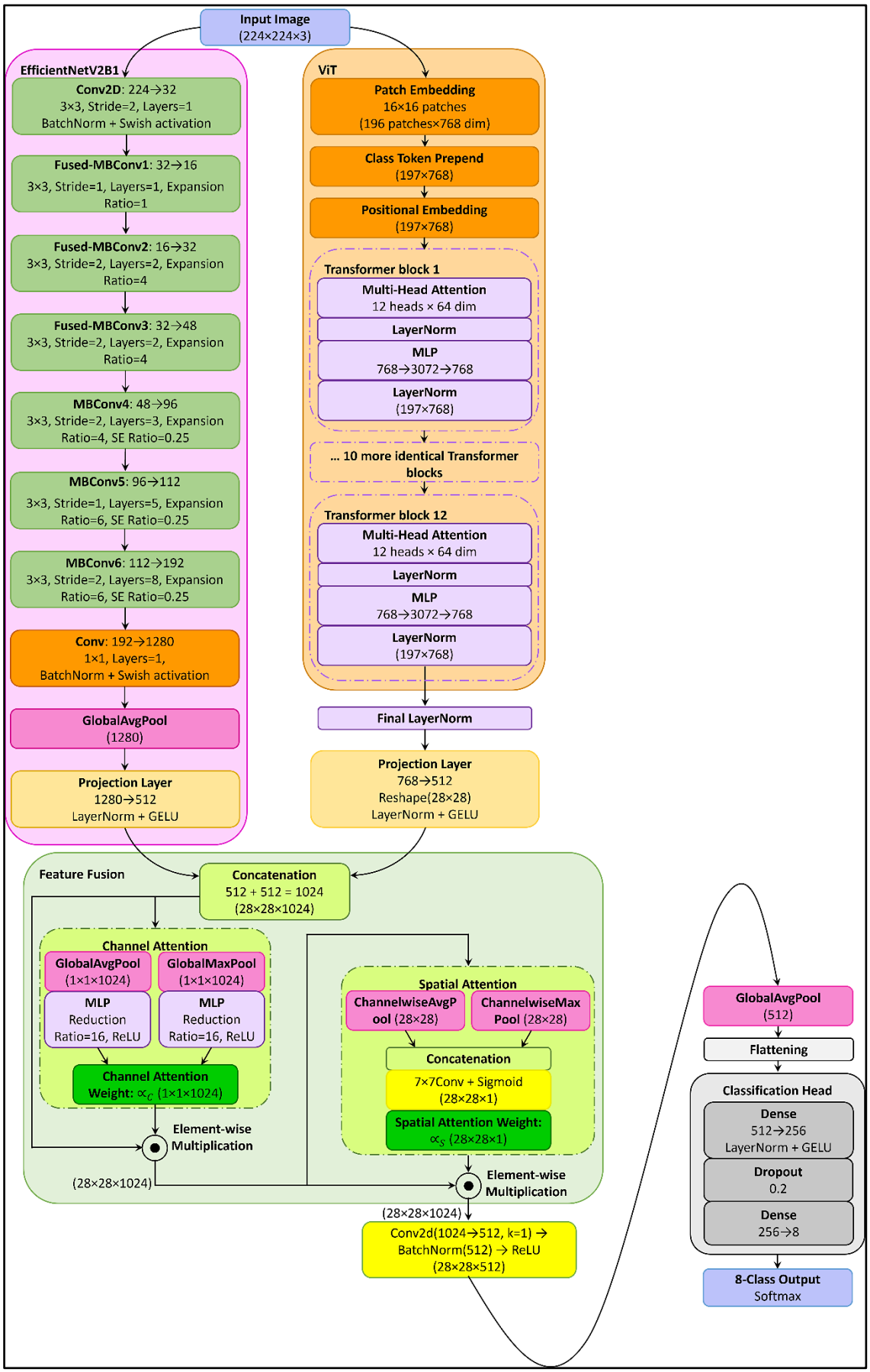

Figure 4.4: Hybrid model of EfficientNetV2B1 and ViT [16]

Chapter 5

RESULTS AND DISCUSSION

5.1 Results of the Proposed Classification Approach Based on LPI Segmentation for Harvested Soybeans

Figure 5.1 illustrates the processing steps for the experimental workflow carried out in this study. A total of 700 grams of harvested soybean seeds were randomly picked from a larger batch. This sample was then split into two portions, maintaining a 70:30 ratio. The larger segment, weighing 490 grams, which is 70% of 700 grams, was considered the training portion, and the remaining 210 grams, which is 30% of 700 grams, was considered for testing the object classification model [27].

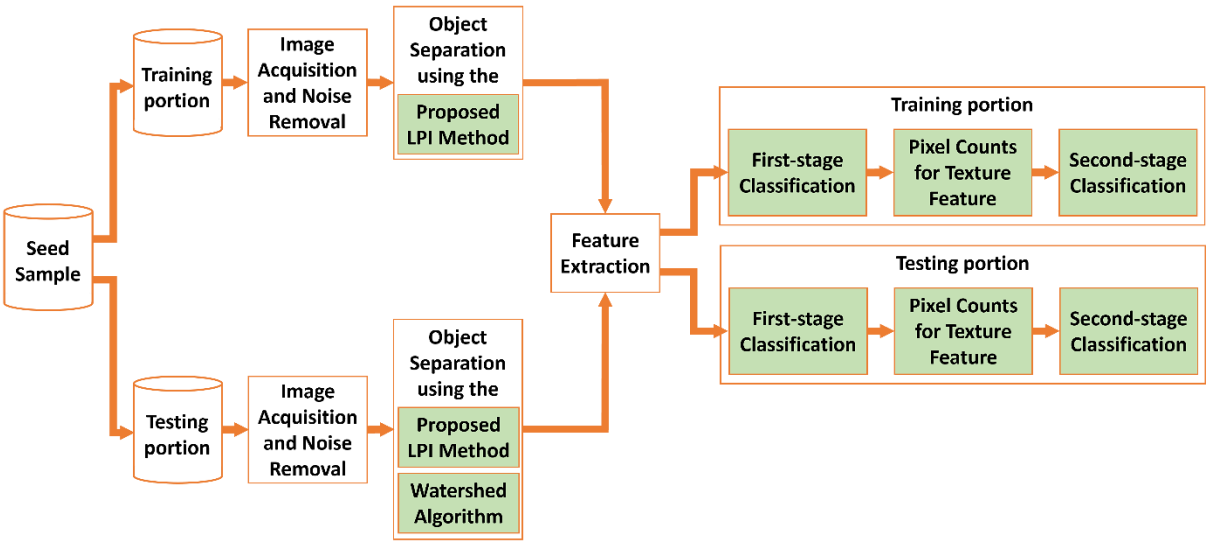

Figure 5.1: Processing steps for classification of harvested seeds

The training portion of the harvested seeds was manually inspected and sorted into eight groups based on their quality and defects, including sound-quality seeds, cracked seeds, split seeds, broken seeds, damaged or discolored or insect-infected seeds, immature or shriveled seeds, organic foreign materials, and inorganic foreign materials. Table 5.1 presents the weights and seed counts for all groups in the training portion of the seed sample, which weighs 490 grams.

Table 5.1: Training portion of the sample with weights and object counts

Group	Weight (in gram)	Seed Count
Sound-quality seeds	175	1522
Cracked seeds	52	485
Split seeds	13	245
Broken seeds	19	332
Damaged or discolored or insect-infected seeds	71	646
Immature or shriveled seeds	11	108
Organic foreign materials	31	279
Inorganic foreign materials	38	276

To manage the available size of the imaging platform, the eight groups of harvested soybeans were divided into smaller chunks, each chunk having an approximate weight of 20 to 40 grams. From the 490 grams of the training portion, smaller chunks were taken based on the group of soybeans or materials. Eight chunks were taken for the sound-quality seeds, three chunks for cracked seeds, two chunks for each of the groups of split seeds, broken seeds, inorganic foreign materials, and organic foreign materials, one chunk for immature or shriveled seeds, and four chunks for the damaged or discolored or insect-infected seeds. For every chunk, images were captured from both the top and bottom sides using the image capture environment designed for this study.

Further, the 210 grams of the testing portion was not sorted or grouped but divided into smaller chunks of weight 42.5 grams. This way, the testing portion was divided into five chunks. For each of these, images were taken from both the top and bottom sides using the imaging environment created for this research. Figure 5.2 shows an example of these images showing the top side of one testing chunk.

5.2.1 Separation of Physically Touching Objects using LPI

For each training portion, either the top or bottom side image was picked at random. The selected image was then processed using the LPI method to separate the harvested seeds in the image.

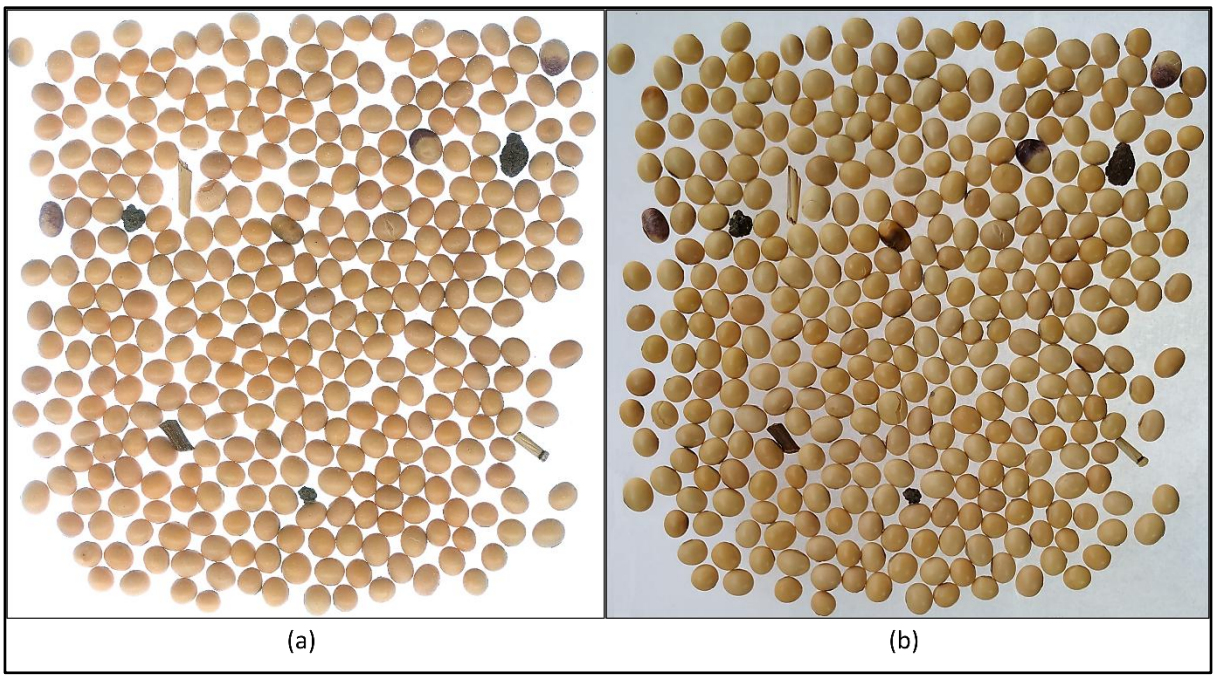

Figure 5.2: Images of a soybean seed sample showing (a) Bottom side, and (b) Top side

About the testing portion, either the top or bottom side image was picked at random for every testing chunk. These images were processed independently using both the watershed algorithm [9] and the LPI method to create black-and-white images in which the objects appear separated. Figure 5.3(a) and (b) show the results from the watershed method and the LPI method, respectively, for the bottom side image in Figure 5.2(a).

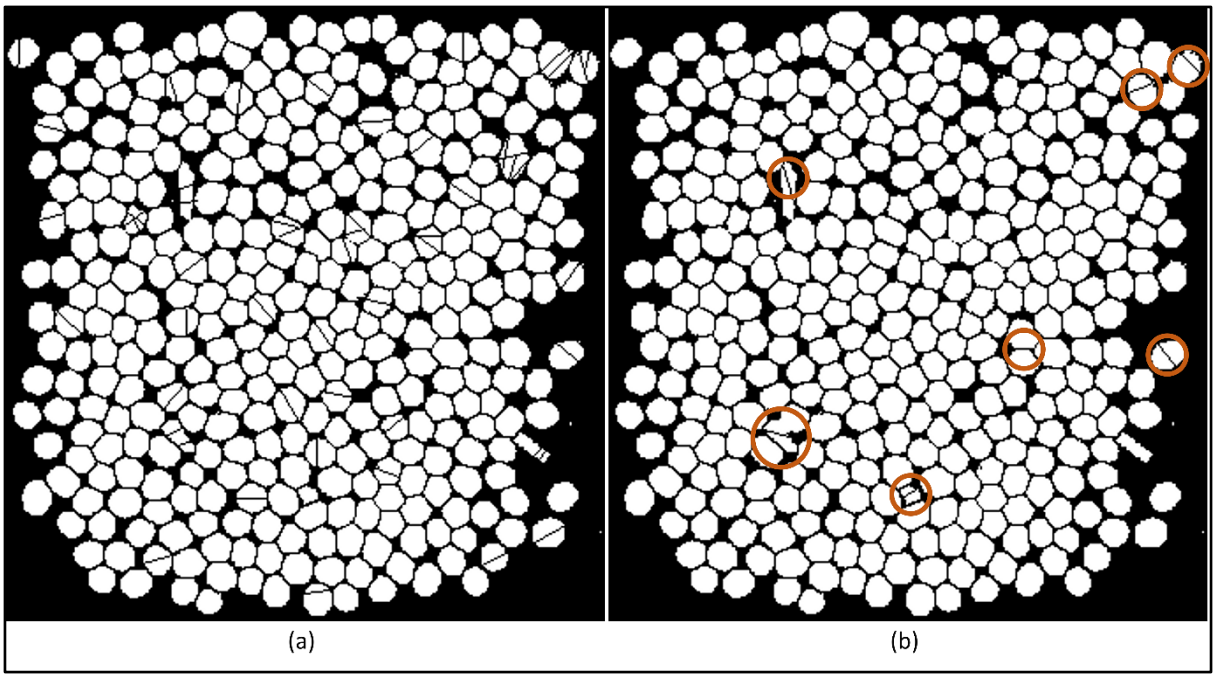

Figure 5.3: Separation of physically touching seed sample objects by (a) Watershed algorithm, and (b) LPI (Orange circle points out seeds that could not be separated or divided into pieces by LPI)

As shown in Figure 5.3(b), the LPI method either divides fewer harvested seeds into pieces or does not separate physically touching seeds due to the non-capturing of their valid V-groove pairs. This means that the difference in the number of harvested seeds detected, compared to the actual count, is much smaller with the LPI method compared to the watershed method. Table 5.2 shows the exact number of harvested seeds, and the counts obtained using both methods for the testing portion of the sample.

According to Table 5.2, the mean difference in harvested seed count is only about 1.93% with the LPI method, whereas the watershed method shows a higher difference of approximately 17.76%. This difference in seed count can negatively impact how accurately the system can classify the soybeans.

Table 5.2: Count of harvested seeds detected by the watershed method and the proposed LPI method for the testing portion of the seed sample

Chunk number	**Actual number of seeds**	**Seeds detected by the watershed method**		**Seeds detected by the LPI method**	
		Count	**% Difference**	**Count**	**% Difference**
1	338	398	17.75	344	1.78
2	342	407	19.01	350	2.34
3	342	403	17.84	348	1.75
4	339	401	18.29	346	2.06
5	346	401	15.90	352	1.73
Mean % Difference =		-	**17.76%**	-	**1.93%**

5.2.2 Classification of Objects Detected using LPI

The weighted KNN (W-KNN) [28], bagged decision trees (B-D-Trees) [29], and cubic SVM (C-SVM) [30] were trained for the classification of harvested seeds. For this purpose, the training portion images processed through the proposed LPI method were utilized to extract the top and bottom side images of every seed. The binary extraction template and the top and bottom side images of every seed were passed to the feature extraction block for extracting geometric and color features. These features were then fed to the above classifiers along with their labels for the classification. The classification performance was checked using the confusion matrix of these classifiers. In this evaluation, the number of correctly identified objects (true positives) was used as the accuracy metric.

During the experimentation, the authors observed that the seeds belonging to the groups of cracked seeds and sound-quality seeds had similar geometry and color features, which decreased the classification accuracy. To address the issue, the initial classification was considered the first-stage classification, and in this classification, cracked seeds and sound-quality seeds were considered in the same group. Table 5.3 shows these true positives for each classifier. It is clear from the table that both B-D-Trees and W-KNN slightly outperform the C-SVM in terms of accuracy.

Table 5.3: True positive values for the first-stage classification

Group	**True positives (in %)**								
	Training portion			Testing portion					
	LPI			LPI			Watershed		
	C-SVM	B-D-Trees	W-KNN	C-SVM	B-D-Trees	W-KNN	C-SVM	B-D-Trees	W-KNN
Cracked or sound-quality seeds	94	94	95	93	94	94	74	73	73
Split seeds	81	89	87	79	87	87	40	45	43
Broken seeds	92	94	95	89	93	92	50	52	52
Damaged or discolored or insect-infected seeds	93	95	96	92	94	95	61	62	60
Immature or shriveled seeds	88	91	92	89	91	92	68	70	71
Organic foreign materials	89	92	93	88	92	93	31	32	34
Inorganic foreign materials	91	94	96	90	93	95	68	69	70

Likewise, for the classification of the testing portion, two separate binary images are created for each chunk image: one using the watershed method and the other using the proposed LPI method. These two images were handled separately. In each one, the distinct seed regions were

detected. Then, every detected seed sample object and its top and bottom side images were processed to extract their features. These features were sent to the classifiers trained in the last step. After classification, the true positives from the confusion matrix were recorded and averaged over five different testing chunks for each seed group. The results in Table 5.3 indicate that the classifiers perform noticeably better when using images processed with the LPI method compared to the watershed method. This increase in accuracy is due to the LPI method being more effective at separating harvested seeds.

Further, the second-stage classification for the training portion was implemented to classify only the seeds identified as cracked or sound-quality seeds in the first-stage classification. For this purpose, the harvested seeds from this group of the first stage were processed for edge detection. The count of pixels of the detected edges was used as a feature, indicating the magnitude of texture on the seed surface to classify the seeds in the second-stage classification. The classifiers Fine Gaussian Support Vector Machine (F-G-SVM) [30], Logistic Regression (L-Regression) [31], and Coarse Decision Tree (CO-Tree) [32] were trained on this feature for binary classification. The classification performance was measured in terms of the true positive accuracy of the classifiers for each group. As illustrated in Table 5.4, all classifiers under investigation achieve nearly the same accuracy.

Table 5.4: True positive values for the second-stage classification

Group	**True positives (in %)**								
	Training portion			Testing portion					
	LPI			LPI			Watershed		
	F-G-SVM	L-Regression	CO-Tree	F-G-SVM	L-Regression	CO-Tree	F-G-SVM	L-Regression	CO-Tree
Sound-quality seeds	94	95	95	94	96	95	74	76	72
Cracked seeds	69	71	70	75	78	78	52	55	52

Similarly, in the second-stage classification, the F-G-SVM, L-Regression, and CO-Tree classifiers were tested on the testing portion of the seed sample. As before, binary images produced using both the watershed method and the LPI method were tested separately. In each case, the pixel count and the label of the object were provided to the classifiers that were trained during the second stage. The true positive values from each classifier's confusion matrix were then averaged over five testing portion chunks for every group. According to Table 5.4, the classifiers performed about 15-20% better when using the LPI-based images compared to the watershed method. Notably, for sound-quality seeds, the classifiers achieved an average accuracy between 94% - 96%, which is a strong result.

To calculate the final accuracy for identifying cracked seeds and sound-quality seeds, the accuracy from the second stage was adjusted based on the results from the first stage. For this, the classification accuracy of the W-KNN from the first stage (testing portion, Table 5.3) and the accuracy of the L-Regression from the second stage (testing portion, Table 5.4) were used. These were combined to determine the overall true positive accuracy across all eight seed

groups. The results are shown in Table 5.5 for both the watershed method and the LPI approach. As shown in Table 5.5, the proposed method detected sound-quality seeds with an accuracy of 90%, which is much higher than the accuracy achieved using the watershed method. This suggests that the automated seed detection system, when built using the proposed LPI method, can grade harvested soybean seeds with an overall accuracy close to 90%.

Table 5.5: Normalized true positive accuracy values for the testing portion

Group	**True positives (in %)**	
	W-KNN with L-Regression	
	LPI method	Watershed method
Sound-quality seeds	90	55
Cracked seeds	73	40
Split seeds	87	43
Broken seeds	92	52
Damaged or discolored or insect-infected seeds	95	60
Immature or shriveled seeds	92	71
Organic foreign materials	93	34
Inorganic foreign materials	95	70

5.2 Results of the Proposed Hybrid Neural Network for Classifying Black-Eyed Peas

5.3.1 Sample Seed Detection using SAM

To assess how well SAM performed with the object-retrieving module, we also tested it alongside a well-known method for separating individual objects, known as the watershed algorithm.

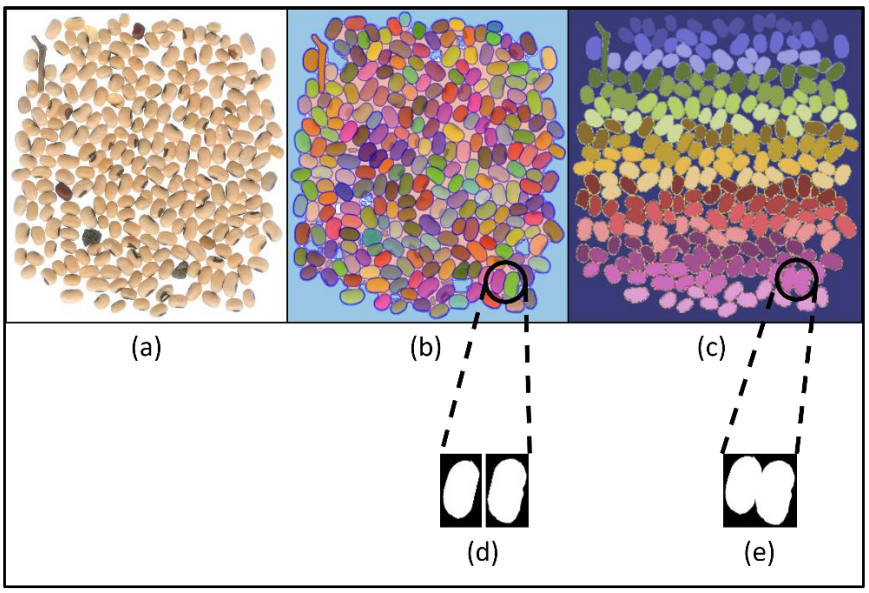

Figure 5.4: (a) Clean image of a harvested black-eyed pea sample, (b) After segmentation by SAM, (c) After segmentation by the Watershed algorithm, (d) Binary extraction templates by SAM, (e) Binary extraction templates by the Watershed algorithm

Figure 5.4 shows the output of both methods. SAM performed effectively and accurately in separating almost all the sample seeds in the image. It was true even when the seeds varied in color, shape, or size, were close together, or were arranged randomly. SAM achieved near-perfect accuracy in this assessment.

5.3.2 Classification of Detected Sample Seeds using Hybrid Neural Network

In the present study, we tested five different hybrid models by combining some well-known transfer-learning-based deep neural networks. The objective was to build a well-trained and accurate model that requires optimal computing power. The hybrid neural networks we used include ResNet50-DenseNet121, InceptionV3-MobileNetV2, EfficientNetV2B1-ViT, VGG16-ResNet18, and MobileNetV2-ShuffleNetV2. We ran these models using Python on our test image dataset of harvested black-eyed peas, and the models were unoptimized for our input dataset in these initial runs. Table 5.6 shows how all five models performed across eight different categories. Among them, the EfficientNetV2B1 combined with ViT gave the best results. Using transfer learning with hybrid models and CBAM attention mechanism with feature fusion technique helped boost the overall performance.

Table 5.6: Overall performance of the hybrid neural networks under investigation

Hybrid Neural network	Inference Time (ms)	Parameters (M)	Recall (%)	F1-Score (%)	Precision (%)	Accuracy (%)
ResNet50 - DenseNet121	49	50.14	92.97	92.94	93.23	93.56
InceptionV3 - MobileNetV2	36	29.23	92.04	92.01	92.42	92.47
EfficientNetV2B1 - ViT	52	56.29	94.58	94.53	94.61	94.60
VGG16 - ResNet18	72	149.76	92.37	92.38	92.65	92.93
MobileNetV2 - ShuffleNetV2	23	13.94	90.97	91.05	91.38	91.64

To improve the performance of the proposed EfficientNetV2B1-ViT model on harvested black-eyed peas, we adjusted its hyperparameters. Table 5.7 enlists the specific hyperparameters and their chosen values used during this process. For tuning, we used the RandomSearch method available in the KerasTuner directory of Python. Instead of testing every possible value, RandomSearch selects combinations randomly, which helps it quickly test a wide variety of options [33]. We ran 50 different trials, with each trial training the model for 40 epochs using different sets of hyperparameters. The optimized best hybrid neural network was then trained on the training portion of the image data and assessed against the validation data to determine its effectiveness. From Figure 5.5, the neural network began with good accuracy because of

the effective tuning techniques previously applied. It reached high accuracy in less than 30 rounds of training.

Table 5.7: Parameters and their values for the optimization of the hybrid neural networks

Parameters	Configuration Options
Pooling2D – height and width of a pool	[1, 2, 3]
Dropout	[0.2, 0.3]
Batch size	[16, 32, 64]
Learning rate	[0.001, 0.01, 0.1]
Optimizer	[Adam, SGD]

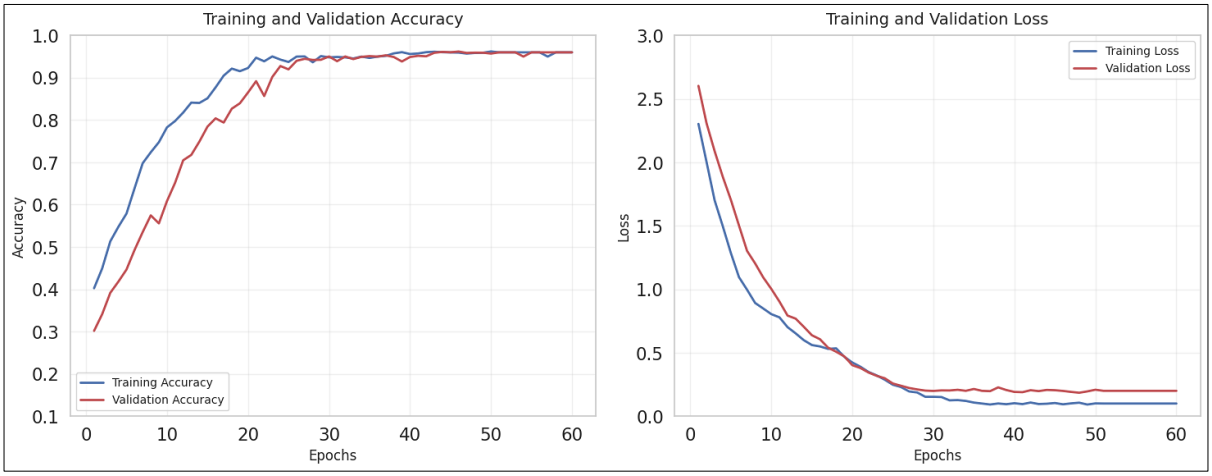

Figure 5.5: Training and validation accuracy and loss trends for the optimized EfficientNetV2B1-ViT Model [16]

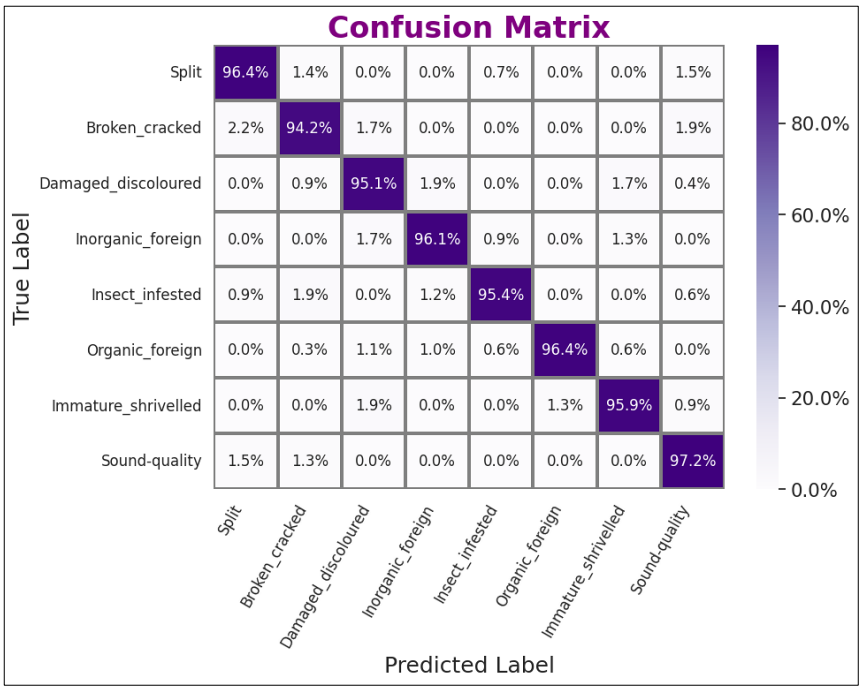

Figure 5.6: Confusion matrix of the optimized EfficientNetV2B1+ViT model on the testing image data [16]

Next, the hybrid neural network was evaluated using new data it hadn't seen before to check how well it could classify the images. As shown in the confusion matrix in Figure 5.6, the model achieved over 94% accuracy in every class. It performed best on sound-quality seeds, with an accuracy of slightly more than 97%, while the minimum accuracy was nearly 94% for the broken and cracked category. These results show that the model was able to clearly distinguish between the various classes and exhibited steady performance overall.

5.3.3 Performance Comparison Between the Existing and Proposed Methods

The comparison table, Table 5.8, highlights the strengths of the proposed method, EfficientNetV2B1+ViT, as it outperforms others in object classification across eight categories. This is a significant step from other methods, which have been limited to classifying only three, four, or five categories.

Table 5.8: Classification accuracies of the existing methods and the proposed method

Method		**Categories of objects in sample**	**Accuracy (%)**
Existing Methods	Principal Component Analysis + Partial Least Squares-Discriminant Analysis (PLS-DA)/ SVM/ KNN (2021 [40])	3	93.90
	Mask R-CNN + S-Net (2022 [43])	5 {Foreign objects not considered}	96.20
	Multi-scale Retinex with color restoration + CNN (2023 [44])	4 {normal, damaged, abnormal, and non-classifiable}	95.63
The Proposed Method	EfficientNetV2B1+ViT	8	95.80

For example, the scheme developed in 2021 could only classify three object types with a classification accuracy of 93.90%. Another method introduced in 2022 was able to classify five categories with a classification accuracy of 96.20%. However, in the real world, the seed samples have more than five categories. The method proposed in 2023 was able to classify four categories (normal, damaged, abnormal, and non-classifiable) with an accuracy of 95.63%. As compared to these methods, the proposed method significantly improved the classification capacity by accommodating eight categories, enabling much more accurate distinctions between various types of objects. This functionality is highly beneficial in real-world classification tasks. By classifying more, this method is better suited for more complex and varied challenges, all with a high accuracy of 95.80%.

Chapter 6

SUMMARY AND CONCLUSION

Crops such as soybeans and black-eyed peas are highly valued for their direct consumption and diverse food industry applications. These legumes serve as raw materials for various food products, including animal feed, oils, biomaterial products, etc. However, using high-quality harvested beans free from contamination by foreign objects for manufacturing can only guarantee end products of standard quality. Food industries conduct quality assessments to ensure the high quality of harvested beans. Evaluating the physical properties of harvested beans is a preliminary step in this process. For this purpose, automated systems are preferred over manual inspection. This research intended to overcome the difficulty of improving the accuracy of the automated systems in classifying harvested beans based on their physical properties. The objective was to improve the preservation of bean features during the image acquisition and bean detection phases.

A novel computing scheme combining two-sided image acquisition and the proposed LPI method was proposed in this work. This novel scheme demonstrates a considerable improvement in the performance of bean detection and classification without compromising the objectives of automation. The performance of the proposed LPI method was consistent across different bean types, including kidney beans, black-eyed peas, and soybeans, and sample weights ranging from 10 to 40 grams at 10-gram intervals. Experimental findings demonstrated that the object detection accuracy for the proposed LPI method was 97%. Also, the suggested approach minimizes the loss of information about the objects, as it separates the physically touching objects with a linear path of the width of the single pixel at the region of contact only. More accurate object detection increases the total performance of the automated system. Further, the classification accuracy classification accuracy was 90% for non-cracked sound-quality soybeans. The accuracy of classifying cracked and split soybeans varied between 70% and 90%. The accuracy for all other categories was above 90%.

However, the experimental results also uncovered some limitations of the proposed method. The performance of the proposed LPI method may decline if a sample image contains a larger number of foreign objects, such as sticks and stems, than the seeds. In such cases, the proposed LPI method must be updated to process diverse non-linear contacts formed by foreign objects.

Moreover, using advanced algorithms like SAM integrated with a custom object retrieval block improved the capability of the system to identify and classify seeds for different sample sizes and inconsistent lighting conditions through instance segmentation. The successful implementation of hybrid models, particularly the EfficientNetV2B1 + ViT, highlights the potential of combining different neural network architectures to improve classification accuracy. EfficientNetV2B1 + ViT model's ability to use local and global features through advanced attention mechanisms resulted in an impressive overall classification accuracy of 95.80%.

Chapter 7

FUTURE WORK

There is scope to extend the proposed work to address the following limitations.

- The pattern-matching windows can be updated to detect V-shaped structures and their corner pixels for non-linear contacts between harvested seeds.
- The LPI algorithm can be improved to separate non-linear contacts of harvested seeds from the sample images.
- A better mechanism can be developed for acquiring two-sided images of clustered seed samples. Enhanced imaging techniques could capture more information about individual seeds in the clustered sample, allowing for a more accurate classification of harvested seeds.
- Also, the proposed methodology may be tested across diverse types of seeds.

PUBLICATIONS

1. S. Sonawane and B. K. Mohanty, “An improved image processing scheme for automatic detection of harvested soybean seeds,” *Journal of Food Measurement & Characterization*, vol. 15, no. 6, pp. 5607–5621, Aug. 2021, doi: https://doi.org/10.1007/s11694-021-01124-0. (SCIE IF: 2.9)
2. S. Sonawane and S. Kurumbanshi, “Robust Classification of Black-Eyed Peas Based on Segment Anything Model and Transfer Learning,” *Journal of Information Systems Engineering and Management*, Vol. 10, no. 21s, pp. 66–80, March 2025, doi: https://doi.org/10.52783/jisem.v10i21s.3290. (SCOPUS indexed)
3. S. Sonawane and B. K. Mohanty, “Detection of harvested beans from a cluster sample,” presented at the 7th International Conference on Computational Optimization, Modelling and Simulation: Recent Trends and Challenges organized by Rungta College of Engineering & Technology, Bhilai, India and sponsored by AICTE, March 25 – 26, 2022, Paper no. 2587.

REFERENCES

[1] S. K. Jensen et al., "Bean and pea-based protein in broiler diets: Performance, feed conversion efficiency and intestinal morphology," Poult. Sci., vol. 102, no. 4, pp. 102287-102296, 2023. doi: 10.1016/j.psj.2023.01.004

[2] Food and Agriculture Organization, "Legume contributions to sustainable livestock production systems," FAO Animal Production and Health Paper No. 189, Rome, Italy, 2023. [Online]. Available: https://www.fao.org/documents/card/en/c/cc3845en

[3] U.S. Department of Agriculture, Foreign Agricultural Service, "Biofuels Annual," GAIN Report BR2024-0022, Brasilia, Brazil, 2024. [Online]. Available: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Biofuels%20Annual_Brasilia_Brazil_BR2024-0022. Accessed: May 23, 2025.

[4] "Post-Harvest Profile of soybean." [Online]. Available: https://www.ams.usda.gov/sites/default/files/media/SoybeanStandards.pdf. Accessed: Apr. 2025.

[5] "Post-Harvest Profile of cowpeas." [Online]. Available: https://www.ams.usda.gov/sites/default/files/media/WholeDryPeas.pdf. Accessed: Apr. 2025.

[6] J. G. A. Barbedo, "Counting Clustered Soybean Seeds," *2012 12th International Conference on Computational Science and Its Applications*, Salvador, Brazil, 2012, pp. 142-145, doi: 10.1109/ICCSA.2012.35.

[7] A. Dell'Aquila, "Computerised seed imaging: a new tool to evaluate germination quality," *Communications in Biometry Crop Science*, vol. 1, no. 1, pp. 20–31, 2006.

[8] M. Wiwart, E. Suchowilska, W. Lajszner, and Ł. Graban, "Identification of hybrids of spelt and wheat and their parental forms using shape and color descriptors," *Computers and Electronics in Agriculture*, vol. 83, pp. 68–76, Feb. 2012, doi: 10.1016/j.compag.2012.01.015.

[9] M. A. Momin, K. Yamamoto, M. Miyamoto, N. Kondo, and T. Grift, "Machine vision based soybean quality evaluation," *Computers and Electronics in Agriculture*, vol. 140, pp. 452–460, Jul. 2017, doi: 10.1016/j.compag.2017.06.023.

[10] K. Sendin, M. Manley, and P. J. Williams, "Classification of white maize defects with multispectral imaging," *Food Chemistry*, vol. 243, pp. 311–318, Sep. 2017, doi: 10.1016/j.foodchem.2017.09.133.

[11] M. Charytanowicz, P. Kulczycki, P. A. Kowalski, S. Łukasik, and R. Czabak-Garbacz, "An evaluation of utilizing geometric features for wheat grain classification using X-ray images," *Computers and Electronics in Agriculture*, vol. 144, pp. 260–268, Dec. 2017, doi: 10.1016/j.compag.2017.12.004.

[12] N. Ansari, S. S. Ratri, A. Jahan, M. Ashik-E-Rabbani, and A. Rahman, “Inspection of paddy seed varietal purity using machine vision and multivariate analysis,” *Journal of Agriculture and Food Research*, vol. 3, p. 100109, Jan. 2021, doi: 10.1016/j.jafr.2021.100109.

[13] Z. Huang *et al.*, “Deep learning based soybean seed classification,” *Computers and Electronics in Agriculture*, vol. 202, p. 107393, Oct. 2022, doi: 10.1016/j.compag.2022.107393.

[14] W. Lin, L. Shu, W. Zhong, W. Lu, D. Ma, and Y. Meng, “Online classification of soybean seeds based on deep learning,” *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106434, May 2023, doi: 10.1016/j.engappai.2023.106434.

[15] Z. Wang, S. Fan, T. An, C. Zhang, L. Chen, and W. Huang, “Detection of Insect-Damaged maize seed using hyperspectral imaging and hybrid 1D-CNN-BILSTM model,” *Infrared Physics & Technology*, vol. 137, p. 105208, Feb. 2024, doi: 10.1016/j.infrared.2024.105208.

[16] S. A. Sonawane, and S. Kurumbanshi, “Robust classification of Black-Eyed peas based on segment anything model and transfer learning,” *Journal of Information Systems Engineering & Management*, vol. 10, no. 21s, pp. 66–80, Mar. 2025, doi: 10.52783/jisem.v10i21s.3290.

[17] X. Sun, X. Li, B. Tan, J. Gao, L. Wang, and S. Xiong, “Integrating OTSU Thresholding and Random Forest for Land Use/Land Cover (LULC) classification and seasonal analysis of water and Snow/Ice,” *Remote Sensing*, vol. 17, no. 5, p. 797, Feb. 2025, doi: 10.3390/rs17050797.

[18] A. Kirillov *et al.*, “Segment anything,” *arXiv (Cornell University)*, Jan. 2023, doi: 10.48550/arxiv.2304.02643.

[19] P. Dhariwal and A. Nichol, “Diffusion models beat GANs on image synthesis,” *arXiv (Cornell University)*, Jan. 2021, doi: 10.48550/arxiv.2105.05233.

[20] H. Pinckaers, B. Van Ginneken, and G. Litjens, “Streaming convolutional neural networks for End-to-End learning with Multi-Megapixel images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 3, pp. 1581–1590, Aug. 2020, doi: 10.1109/tpami.2020.3019563.

[21] M. E. Sakka, M. Ivanovici, L. Chaari, and J. Mothe, “A review of CNN applications in smart agriculture using multimodal data,” *Sensors*, vol. 25, no. 2, p. 472, Jan. 2025, doi: 10.3390/s25020472.

[22] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” *arXiv (Cornell University)*, Jan. 2019, doi: 10.48550/arxiv.1905.11946.

[23] M. Tan and Q. V. Le, “EfficientNetV2: Smaller models and faster training,” *arXiv (Cornell University)*, Jan. 2021, doi: 10.48550/arxiv.2104.00298.

[24] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv (Cornell University)*, Jan. 2020, doi: 10.48550/arxiv.2010.11929.

[25] S. Lei, C. Xia, Z. Li, X. Li and T. Wang, "HNN: A Novel Model to Study the Intrusion Detection Based on Multi-Feature Correlation and Temporal-Spatial Analysis," in *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 4, pp. 3257-3274, 1 Oct.-Dec. 2021, doi: 10.1109/TNSE.2021.3109644.

[26] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” *arXiv (Cornell University)*, Jan. 2018, doi: 10.48550/arxiv.1807.06521.

[27] Y. Xu and R. Goodacre, “On splitting training and validation set: A comparative study of Cross-Validation, Bootstrap and systematic sampling for estimating the generalization performance of supervised learning,” *Journal of Analysis and Testing*, vol. 2, no. 3, pp. 249–262, Jul. 2018, doi: 10.1007/s41664-018-0068-2.

[28] S. S. Mullick, S. Datta and S. Das, "Adaptive Learning-Based k -Nearest Neighbor Classifiers With Resilience to Class Imbalance," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5713-5725, Nov. 2018, doi: 10.1109/TNNLS.2018.2812279.

[29] O. Al–Khadher, A. Mukhtaruddin, F. R. Hashim, M. M. Azizan, H. Mamat and N. Saleh, "Non-intrusive load monitoring based on bagging decision trees and the selective features for commercial building loads," *2022 IEEE International Conference on Power and Energy (PECon)*, Langkawi, Kedah, Malaysia, 2022, pp. 293-298, doi: 10.1109/PECon54459.2022.9988836.

[30] A.-K. Al-Khowarizmi, I. P. Sari, and H. Maulana, “Optimization of support vector machine with cubic kernel function to detect cyberbullying in social networks,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 2, p. 329, Mar. 2024, doi: 10.12928/telkomnika.v22i2.25437.

[31] F. Abramovich, V. Grinshtein and T. Levy, "Multiclass Classification by Sparse Multinomial Logistic Regression," in *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4637-4646, July 2021, doi: 10.1109/TIT.2021.3075137.

[32] Y. Xie, C. Zhu, R. Hu, and Z. Zhu, “A Coarse-to-Fine Approach for Intelligent Logging Lithology Identification with Extremely Randomized Trees,” *Mathematical Geosciences*, vol. 53, no. 5, pp. 859–876, Aug. 2020, doi: 10.1007/s11004-020-09885-y.

[33] Keras tuner. https://keras.io/keras_tuner/. Accessed in November 2024.